Year : 2023

Quantitative genetic variation in symbiotic fungi and its effect on fungal gene expression

Cruz Corella Joaquim

Cruz Corella Joaquim, 2023, Quantitative genetic variation in symbiotic fungi and its effect on fungal gene expression

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_27A1E96100873

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Department of Ecology and Evolution

Quantitative genetic variation in symbiotic fungi and its effect on fungal gene expression

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

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

par

Joaquim Cruz Corella

Biologiste diplômé d'un MSc en Microbiologie Avancée
University of Barcelona, Spain

Jury

Prof. Christian Fankhauser, Président
Prof. Ian R. Sanders, Directeur de thèse
Prof. Marc Robinson-Rechavi, Expert interne
Prof. Daniel Croll, Expert externe

Lausanne - 2023

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Department of Ecology and Evolution

Quantitative genetic variation in symbiotic fungi and its effect on fungal gene expression

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

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

par

Joaquim Cruz Corella

Biologiste diplômé d'un MSc en Microbiologie Avancée
University of Barcelona, Spain

Jury

Prof. Christian Fankhauser, Président
Prof. Ian R. Sanders, Directeur de thèse
Prof. Marc Robinson-Rechavi, Expert interne
Prof. Daniel Croll, Expert externe

Lausanne - 2023

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.	Christian	**Fankhauser**
Directeur·trice de thèse	Monsieur	Prof.	Ian R.	**Sanders**
Expert·e·s	Monsieur	Prof.	Marc	**Robinson-Rechavi**
	Monsieur	Prof.	Daniel	**Croll**

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

Joaquim Cruz Corella

Master - Màster universitari en Microbiologia avançada, Universitat de Barcelona, Espagne

intitulée

Quantitative genetic variation in symbiotic fungi and its effect on fungal gene expression

Lausanne, le 23 juin 2023

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Christian Fankhauser

A la meva família

Table of Contents

Acknowledgements

I would need several pages to properly thank everyone who has significantly contributed in the making of this PhD thesis, in one way or another. I will try my best to concisely acknowledge them here.

Firstly, I would like to thank Prof. Ian R. Sanders who gave me what will most likely be the greatest opportunity in my scientific career. Being one of his students has been a privilege. He gave me freedom to discover the fascinating world of AMF at my own pace. He always asked the right questions and provided the best feedback. His guidance, support and understanding through these years have been crucial for me, to an extent that is difficult to put in writing.

I would like to thank my two experts Prof. Marc Robinson-Rechavi and Prof. Daniel Croll for kindly accepting to read and review the work in this thesis, as well as the president of the jury Prof. Christian Fankhauser.

I would like to particularly thank Frédéric Masclaux, Réjane Seiler, Soon-Jae Lee, Ivan Mateus, Ricardo Peña, Cristian Rincón, Romain Savary, Tania Wyss, Lucas Villard, Nicolo Tartini, Anurag Chaturvedi, Chanz Robbins and the whole of the current Sanders group. A quick mention to Alia Rodriguez for her support and the rest of the team in Colombia. Also, I would like to thank the many nice people from the DEE that I've met throughout the years that I cannot list all here.

This PhD adventure started while I was a teaching assistant at the University of Lleida. I would like to thank Dra. Ester Vilaprinyó and Prof. Rui Alves for their trust, support, and motivation; without them I wouldn't have considered pursuing a PhD abroad in the first place.

I would like to thank all my closest friends who always had kind and warming words in moments of doubts and difficulties. Anaïs Hermoso, Sílvia Freixes, Carlota Rodriguez, Mònica Boladeres (who draw the awesome and colourful illustration of AMF in the back cover of the thesis), Nuria Pijuan, and the rest of the good friends in Balaguer. Elisa Battikh, Xènia Ollé and the rest of the Catalan friends I've made in Switzerland. Violeta Echeverría, my first and closest friend in Lausanne, and the rest of the Spanish family. Rebecca Leghziel and Manuel Cherep, my two incredibly smart and cherished flatmates. Oliver Maldonado and Alex Cabré who know the real magic of friendship. Sandra Bosch, Daniel Isábal and the rest of the friends from biotech. Rodrigo Fraile, Javi García, Carlos Lahera and Jacobo Bustamante, of course. Imanol Cabana and Sophie Bally, whose friendship has been truly uplifting.

Lastly, I would like to enormously thank my mum, my dad, and my brother whose love, generosity and support are beyond invaluable; and to Judit Llobera, for supporting me during all these years.

I am incredibly grateful to you all.

Abstract

Arbuscular mycorrhizal fungi (AMF) are wide-spread soil microorganisms that form a symbiotic relationship with most terrestrial plant species. AMF colonize plants roots, where they exchange inorganic soil nutrients for plant carbohydrates and lipids. This mycorrhizal association dates to more than 400 Myr ago and it has been hypothesized that it may have enabled the transition of plants from aquatic to terrestrial ecosystems. AMF have been repeatedly shown to significantly influence plants growth, plant community structure, plants resistance to various biotic (pathogens) and abiotic stresses (drought, pollutants, salinity, etc.). Therefore, AMF potentially represent a major tool to improve current agricultural practices and increase crop yields. However, the variability observed in these beneficial effects is very large and, to this date, unpredictable. Because different AMF isolates have differential effects on plants, it has become paramount to understand the molecular differences among AMF isolates. Several genetic studies have shown that extensive variation exists within the model AMF species *Rhizophagus irregularis*. *R. irregularis* isolates can be homokaryon or dikaryon, harbouring either one or two major populations of nucleus genotypes. In this thesis, I investigated some of the genetic and epigenetic differences between homokaryon and dikaryon *R. irregularis* isolates. In chapter 2, I studied the presence of the DNA methylation mark N6-methyldeoxyadenine (6mA) in the genome of *R. irregularis* and I found that it is most likely a functional epigenetic mark, and that genetically indistinguishable homokaryon isolates differed in their epigenome. In chapter 3, I investigated the generation of quantitative genetic variation during the clonal reproduction of a dikaryon AMF; and its potential impact on gene expression. I found that the two nucleus genotypes could be found at different ratios in a set of clonally produced offspring, and that the nuclear ratios directly impacted the transcriptome. Lastly, in chapter 4, I continued to investigate the contribution to trait variation of epigenetics in homokaryons, and variable nucleus genotype ratios in dikaryons. I studied the potential variation during the critical pre-symbiotic phase of spore germination, and I found that dikaryons generated greater plasticity and variation in molecular responses to changes in the environment than homokaryons. This thesis showed that both factors, epigenetics and variable nuclear ratios in dikaryons, are most likely contributing to the still unexplained trait variation in AMF and variable effects on plants.

Résumé

Les champignons mycorhiziens arbusculaires (AMF) sont des micro-organismes du sol très répandus qui entretiennent une relation symbiotique avec la plupart des espèces végétales terrestres. Les AMF colonisent les racines des plantes, où ils échangent les nutriments inorganiques du sol contre des glucides et des lipides. Cette association mycorhizienne existe depuis plus de 400 millions d'années et l'hypothèse a été émise qu'elle aurait permis la transition des plantes des écosystèmes aquatiques aux écosystèmes terrestres. Il a été démontré à plusieurs reprises que les AMF influencent de manière significative la croissance des plantes, la structure des communautés végétales, et la résistance des plantes à divers stress biotiques (ex., pathogènes) et/ou abiotiques (ex., sécheresse, polluants, salinité). Par conséquent, les AMF représentent potentiellement un outil essentiel pour améliorer nos pratiques agricoles courantes pour augmenter les rendements des cultures. Cependant, la variabilité observée dans les effets bénéfiques résultant de l'application d'AMF sur les plantes est très importante et, à ce jour, imprévisible. Étant donné que différents isolats de la même espèce d'AMF ont des effets différentiels sur les plantes, il est devenu primordial de comprendre les différences moléculaires entre ces isolats. Plusieurs études génétiques ont montré qu'il existe une grande variation dans les isolats de l'espèce modèle de AMF, *Rhizophagus irregularis*. Les isolats de *R. irregularis* peuvent être homocaryon ou dikaryon, abritant soit une ou deux populations majeures de génotypes de nucleus, respectivement. Dans cette thèse, j'ai étudié certaines des différences génétiques et épigénétiques entre les isolats homokaryon et dikaryon de *R. irregularis*. Dans le chapitre 2, j'ai étudié la présence de la marque de méthylation de l'ADN N6-méthyldeoxyadénine (6mA) dans le génome de *R. irregularis*. J'ai trouvé qu'il s'agit très probablement d'une marque épigénétique fonctionnelle, et que les isolats homokaryons génétiquement indiscernables différaient dans leur épigénome. Dans le chapitre 3, j'ai étudié la génération d'une variation génétique quantitative pendant la reproduction clonale d'un AMF dikaryon et son impact potentiel sur l'expression génétique. J'ai découvert que les deux génotypes de noyaux pouvaient se trouver à des ratios différents dans un ensemble de descendants produits par clonage, et que les ratios de noyaux avaient un impact direct sur le transcriptome. Dans le chapitre 4, j'ai continué à étudier la contribution de l'épigénétique à la variation des caractères chez les homocaryons et les rapports variables des génotypes de noyau chez les dikaryons. J'ai étudié la variation potentielle pendant la phase critique pré-symbiotique de la germination des spores, et j'ai découvert que les dikaryons généraient une plus grande plasticité, et une plus grande variation, que les homokaryons, dans les réponses moléculaires aux changements de l'environnement. Cette thèse a montré que deux facteurs, l'épigénétique et les rapports nucléaires variables chez les dikaryons, contribuent très probablement à la variation des traits encore inexpliquée chez les AMF qui peuvent être à l'origine de leurs effets variables sur les plantes.

Declaration

Chapters 2, 3 and 4 containing original experiments and data are written as manuscripts for publication. Chapter 2 was already published in Communications Biology on July 22nd, 2021. Chapter 3 was published in New Phytologist on June 4th, 2021. Chapter 4 has been submitted for publication in New Phytologist on February 22nd, 2023. As is the case in most scientific investigations, the work presented in chapters 2-4 represented collaborations between several colleagues and collaborators. In the case of all three manuscripts, I was jointly responsible for the conception of the studies, collection and analysis of data and writing the manuscripts. The specific contributions of myself (JCC) and each of the co-authors are given below:

Chapter 2: AC, FGM, JCC, CR, IRS conceived the study. AC, JCC, CR, SL, FGM performed data analysis. CR, SL, LM, NG performed experimental work. All authors contributed to writing and editing the manuscript. IRS acquired project funding.

Chapter 3: CR, JCC, and IRS designed experiments; CR, CA, and RS produced ddRAD-seq data on 48 SSSLs; JCC, CA, SJL and RS produced RNA-seq data on 6 dikaryon SSSLs; FM generated ddRAD-seq on 5 dikaryon SSSLs; CR and JCC analysed sequencing data; CR, JCC, SJL, IDM and IRS interpreted data and wrote the manuscript; IRS acquired project funding.

Chapter 4: JCC, SJL and CR conceived the study. JCC, SJL and CR performed the experiments. JCC and SJL analysed the data. JCC, SJL and IRS interpreted data and wrote the manuscript. IRS acquired project funding.

The thesis abstract, the general introduction (**Chapter 1**), and the general discussion (**Chapter 5**) are entirely my own work.

Chapter 1

General introduction

General introduction

Symbiotic associations are pivotal to life on Earth. The evolution of eukaryotic cells was made possible solely by the association of ancient prokaryotic cells with other prokaryotic cells that sheltered within their cytosol. These endosymbionts evolved into what today are acknowledged as essential organelles in the eukaryotic lineages, such as the mitochondria and chloroplasts (Sagan, 1967). Evolution cannot be understood without considering the many and diverse symbiotic interactions that formed throughout the planet since the origin of life.

The term *symbiosis* was coined by the German botanist and mycologist Albert Bernhard Frank in 1877. It refers to the common life of two dissimilar organisms, that can either be a mutually beneficial interaction, or one where one of the partners exploits the other. While the former is typically referred as a "mutualistic" association (e.g., lichens, mycorrhizas, nitrogen-fixing systems, etc.), the latter is referred as a "parasitic" association (e.g., rusts and mildews). Nevertheless, there is consensus among the scientific community that these symbiotic associations can take any form within a continuum between mutualism to parasitism, and that the relationship can even change over the course of a partnership (Ewald, 1987; Bronstein, 1994). It is the persistent fight between the two partners, the symbionts, to gain control over the other through evolutionary innovations, often illustrated as an arms race, that keeps a fragile equilibrium that can lead to what is known as co-evolution (VanValen, 1973). The relation between the symbionts and with the environment shapes the nature of the interaction and its evolutionary fate (Drew et al., 2021).

Among the most ancient plant-microbe associations is the arbuscular mycorrhizal (AM) symbiosis. This is a widespread mutualistic association that is formed by the AM fungi (AMF) and more than 80% of terrestrial plant species (Brundrett, 2004; van der Heijden et al., 2015; Brundrett and Tedersoo, 2018), and dates back to the Devonian, more than 400 Myr ago (Berbee and Taylor, 1993; Remy et al., 1994; Redecker et al., 2000; Strullu-Derrien et al., 2018). It is currently believed that the first establishment of the AM symbiosis was determinant for the plants to transition from aquatic to terrestrial ecosystems (Pirozynski and Malloch, 1975). The AM association is based on a mutually beneficial nutrient exchange that occurs inside plants roots. While the fungus enhances the plant uptake of minerals from the soil, the plant provides its fungal partner with carbon compounds and lipids produced via photosynthesis (Gianinazzi et al., 2010; Wipf et al., 2019). This flux of nutrients is very context-dependent and is highly regulated by both, the fungus and the plant (Kiers et al., 2011).

The AM symbiosis has been known for more than a century (Frank and Trappe, 2005), but it is over the last twenty-five years that research in the field has greatly intensified, partially motivated by its potential to improve agriculture. Food security is a major concern globally as current agricultural practices for food production cannot keep up with human population growth (FAO, 2017). In addition, the effects of global warming, deforestation and, soil and water contamination on agriculture are yet difficult to predict at mid and long-term. In order to reduce the anthropogenic pressure on natural resources, efforts are being made to transition from classical food production and intensified

agriculture to more sustainable systems. Efficiently using natural resources is crucial to achieve sustainability and reduce the human footprint, but might not be sufficient to avoid a scenario of food scarcity if crop yields cannot also be improved significantly. In this context, a well thought out manipulation of the AM symbiosis could provide a chance for potentially fine-tuning and optimizing a plant's acquisition of nutrients and, thus, it could tackle both of these problems at once.

Many publications have described the beneficial effects of AMF on plants and natural ecosystems, including an increase of plant growth and the resistance to various biotic and abiotic stresses (van der Heijden et al., 1998; Rooney et al., 2009; Smith and Smith, 2012). Indeed, there is a huge potential for improving crop yields through the usage of AMF as so-called biofertilizers (Biermann and Linderman, 1983; Rodriguez and Sanders, 2015; Bitterlich et al., 2018). However, the enormous complexity of the AM symbiosis has hindered building effective predictive models for their use in improving plant growth, and poses a serious problem to bring the findings from controlled trials to farming conditions. Complexity comes not only from the large range of environmental factors that are known to heavily influence plant growth in the field, but also from the intrinsic complexity existing at the molecular level within the plant and the fungus, and the many possible interactions among all these factors. In this thesis, I focused on characterizing, at the molecular scale, different sources of variation in AMF that could influence relevant fungal traits by affecting how their own genes are being expressed, with the ultimate goal of understanding whether these molecular mechanisms could have a significant impact on plant growth.

A brief insight into the curious biology of AMF

AMF are a group of the *early diverging fungi* (EDF) that are taxonomically classified as the Glomeromycotina subphylum (Schüβler et al., 2001). They are obligate symbionts which rely on the formation of the symbiosis to survive, and they can be found worldwide in practically every soil type (Brundrett, 2004; Brundrett and Tedersoo, 2018).

AMF grow as an underground filamentous network that associates with plants roots, acting as an extended root system (Friese and Allen, 1991). The network is composed of filaments named *hyphae*. These hyphae are coenocytic, which means that they do not have individual compartments, each containing one nucleus but rather a number of nuclei co-existing in the same cytoplasm. In fact, no single-cell stage is known in the entire AMF life cycle. The cytoplasm, nuclei and organelles are all shared within a continuous filamentous space. AMF are a canonical example of a multinucleated microorganism.

AMF mycelium, known as the *intra-radical mycelium* (IRM), grows inside plants roots. Some of the hyphae develop specialised tree-like structures inside the plant cell wall called arbuscles, where the nutrient exchange between the fungus and the plant takes place (Luginbuehl and Oldroyd, 2017). The fungus also produces round-shaped vesicles inside roots that serve as storage organs (Jabaji-Hare et al., 1984). The *extra-radical mycelium* (ERM) is the part of the mycelium that grows outside plants

roots. The ERM explores the soil, absorbing inorganic nutrients. It also produces in turn a large number of spores which represent fungal reproductive structures (Viera and Glenn, 1990; Marleau et al., 2011).

AMF have historically been considered an ancient asexual eukaryotic lineage, although most information on this topic comes from the study of just one AMF species, *Rhizophagus irregularis* (Sanders, 1999). The claim for asexuality was based on apparent vegetative incompatibility between different AMF isolates (Giovannetti et al., 2003), the lack of any observed sexual structures and of karyogamy. In addition, more recently, a large-scale genomic survey of different AMF isolates from intercontinental locations revealed that thousands of genetic markers were shared among all these isolates (Davison et al., 2015; Savary et al., 2018). This ubiquitous distribution of almost identical genotypes is an expected feature of primarily clonally reproducing species that rarely recombine (Sanders, 2018).

AMF are known to reproduce clonally via large multinucleated spores that germinate to produce new mycelium (Viera and Glenn, 1990). During spore formation a massive migration of nuclei occurs towards the new structures, where they are thought to also multiply by mitosis (Marleau et al., 2011). This process of nuclear migration has been described as stochastic and can generate some variability in the number of nuclei mobilized into each of the many spores of an AMF, although such variability did not significantly affect their capacity to germinate (Marleau et al., 2011; Masclaux et al., 2018).

Recent findings using DNA sequencing technologies have challenged the asexual status of AMF (Bruns et al., 2018). Firstly, traces of genetic recombination were found in some AMF isolates that originated from the same field, which suggested the existence of cryptic sexual reproduction (Croll and Sanders, 2009). Secondly, several meiosis related genes were identified in the AMF genomes, which provided a potential mechanistic explanation for the suspected recombination (Halary et al., 2011; Corradi and Lildhar, 2012). Thirdly, it was recently shown that among a collection of different AMF isolates some were dikaryons, containing a population of two haploid nucleus genotypes (Ropars et al., 2016). In contrast, the vast majority were homokaryons and harboured a homogenous population of haploid nuclei. This was established on the basis of both, an unusually high degree of genetic polymorphism within dikaryon isolates compared to homokaryons and the presence of two different alleles of nucleus specific loci, including a putative MAT -locus that resembled genes involved in mating in other fungal species that undergo sexual reproduction (Ropars et al., 2016; Wyss et al., 2016; Masclaux et al., 2018). AMF dikaryon isolates were hypothesized to originate from the mating of two homokaryons, and it was speculated that the two different nucleus genotypes would only coexist temporarily before they fuse to generate diploid nuclei that undergo meiosis, completing the whole sexual cycle (Ropars et al., 2016; Corradi and Brachmann, 2017).

The exchange of nuclei and organelles between different AMF isolates can occur via *anastomosis*, the fusion of two hyphae (Giovannetti et al., 1999; Chagnon, 2014). Indeed, anastomosis is frequently observed in AMF, even between hyphae of the same isolate. The study of anastomosis between AMF isolates revealed that the frequency of so-called “perfect fusion”, where protoplasmic streaming occurs

within the fusion bridge, was negatively correlated to the genetic distance between a given pair of AMF isolates, suggesting that more distant isolates struggle to form anastomosis, and that vegetative incompatibility may be a gradual process in AMF (Croll et al., 2009). However, the exact compatibility mechanisms regulating under which circumstances anastomosis occurs in AMF, and whether the resulting nuclear exchange is a stable combination that can eventually complete the sexual cycle, are both currently unresolved questions. Thus, despite the compelling possibility of sexual reproduction in AMF, it remains to be empirically demonstrated whether any form of sexual or parasexual reproduction exists in these fungi in addition to the clonal reproduction.

Intra-specific variation in the model AMF species *Rhizophagus irregularis*

Before the rise of molecular techniques phenotypic trait variation in AMF was first documented among different species based on their different spore morphology (Mosse and Bowen, 1968; Gerdemann and Trappe, 1974); and then, based on the presence of polymorphisms in their ribosomal DNA sequences (Sanders et al., 1995); their distinct ability to take up phosphorus and stimulate plant growth (Ravnskov and Jakobsen, 1995; Hart and Reader, 2002). However, the observation of significant trait variation within a single AMF species sparkled the question of whether intra-specific diversity could also be an ecologically important factor capable of influencing plants traits (Munkvold et al., 2004). Several studies addressed this particular question, consistently finding that intra-specific variation in AMF significantly affected plant growth in a wide range of experimental conditions and with different plant species (Koch et al., 2006; Angelard et al., 2010; Ceballos et al., 2019; Peña et al., 2020). It is currently well established in the field of ecology that trait variation within a species can be as extensive as variation across species, and it can significantly affect community structure and ecosystem function (Des Roches et al., 2018). However, at their time, these findings were truly ground-breaking and opened a completely new venture in the research field of mycorrhizal fungi.

Evidence of genetic variation among different AMF isolates of the model species *Rhizophagus irregularis*, formerly miss-assigned to *Glomus intradices*, was first obtained using amplified fragment length polymorphism (AFLP) fingerprinting, and then using a multilocus genotyping approach that targeted simple sequence repeats (SSR) (Koch et al., 2004; Koch et al., 2006; Croll et al., 2008). This intra-specific genetic variation was linked to variable plants responses. However, it was after the publication of the first complete AMF genome by Tisserant *et al.* (2013) that more extensive research studies uncovered the high degree of intra-specific genetic variation found in this AMF species (Tisserant et al., 2013). Most notably, several gene families were found to have suffered large expansions or contractions by comparing whole assembled genomes of six different *R. irregularis* isolates, with up to 50% of their coding sequences being considered accessory (Chen et al., 2018). AMF are versatile symbionts that seem to have little host specificity, and it is critical for their survival to be adaptive to a wide range of rapidly changing environments. Even so, not every AMF exerts the same effects on a given plant host and, thus, some "preferences" could exist between the partners when the symbiosis is first established (Sanders and Fitter, 1992). In line with this reasoning, intra-specific

genetic diversity in AMF is thought to serve this purpose, providing AMF isolates with different toolkits with some core (essential) and some accessory (context-dependent) functionalities.

Trait variation within a species can also arise by other means than genetic variation, e.g., differential epigenetic regulation of the genome. Among the most widespread forms of epigenetic regulation, DNA methylation has been described in many different species across the fungal tree of life. More specifically, N6-methyldeoxyadenine (6mA) and 5-methylcytocine (5mC) were both reported as functional epigenetic marks in the fungal kingdom (Mondo et al., 2017; Bewick et al., 2019). High levels of 6mA were found in a set of EDF species, while very low levels were detected in dikarya. Its presence on gene promoter regions was linked to active transcription. Instead, 5mC was highly correlated with the presence of transposable elements and it is hypothesized to inhibit their activity. Remarkably, despite their ecological importance, no AMF species were included in either of the two studies on DNA methylation in the fungal kingdom.

Given the cryptic nature of the AMF life cycle and their complex multinucleated genetic organization, it remains critical to elucidate by which molecular mechanisms such impressive intra-specific variability arises and how these can truly contribute to AMF functional diversity. In **chapter 2**, we produced a new set of genome assemblies of six *R. irregularis* isolates using long-read sequencing technology, which allowed us to investigate previously uncharacterized large genomic rearrangements and most importantly, the potential existence of 6mA DNA methylation in this species and search for clues to its role in AMF gene regulation.

Generation of new *R. irregularis* single-spore lines

One practical goal of the AMF research is to generate new fungal lines with improved properties for plant growth. The inheritance of genetic material and the emergence of novel traits in new generations of AMF has mostly been studied using two main different culturing techniques that were initially developed to propagate fungal isolates *in vitro*. These techniques differ in the amount of starting fungal material that is transferred into a new plate (to establish a new culture), being either a single spore (single spore culturing) or many spores and hyphae (sub-culturing). Interestingly, sub-culturing of AMF was shown to produce no, or very little, phenotypic and genetic variation among the new cultures, whereas single spore culturing appeared to be much more prone to generate variation (Ehinger et al., 2012).

Croll *et al.* (2009) not only showed that genetic exchange between genetically different *R. irregularis* isolates was possible, but also that single-spore progeny of pairings of different parental isolates (*crossed* lines) significantly differed in some quantitative fungal traits, such as spore and hyphal densities (Croll et al., 2009). At that time, these crossed lines were suspected to be *heterokaryons* (Kuhn et al., 2001; Pawlowska and Taylor, 2004), harbouring a mix of several different nucleus genotypes, originated from the crossing of the two parental isolates. Following this study, Angelard *et al.* (2010) investigated the occurrence of a partial inheritance of those different nucleus genotypes

among newly generated single-spore progeny of a parental *R. irregularis* crossed line (*segregated* lines or, in this thesis, *single spore sibling lines* (SSSLs), as well as their effects on plant growth (Angelard et al., 2010). It was hypothesized that a random mobilization of nuclei into the spores would generate some variability in the proportions at which the different nucleus genotypes were inherited by each individual spore. This could, in turn, affect the fungal phenotype and, ultimately, plant growth. Remarkably, these clonally produced SSSLs not only displayed evidence of differential proportions of nuclear inheritance; the differences observed in rice dry weight among SSSLs treatments were also striking, as they were up to five-fold. Furthermore, posterior studies revealed that these same SSSLs also significantly differed in their quantitative fungal traits and their capacity to colonize plants roots (Angelard and Sanders, 2011; Ehinger et al., 2012).

Despite that some of the *R. irregularis* SSSLs parental lines were initially suspected to be heterokaryons, more recent evidence showed that *R. irregularis* isolates appear to be either homokaryons or dikaryons (Ropars et al., 2016). Given that single spore progeny of these *R. irregularis* parental isolates could only inherit genetic material which was already present in the parent, all SSSLs generated by Angelard *et al.* (2010) must be homokaryon or dikaryon too, and should carry the same exact alleles. Interestingly, the unequal inheritance of nucleus genotypes among SSSLs was also observed in dikaryons by Masclaux *et al.* (2019), that determined the allele frequencies of the nucleus genotype specific locus *bg112* using amplicon sequencing (Masclaux et al., 2018). This form of quantitative genetic variation was again hypothesized to contribute substantially to trait variation among these dikaryon sibling lines. In **chapter 3**, we studied what was the effect of the different nucleus genotype ratios in dikaryon SSSLs transcriptomes. We performed double digest restriction-site associated DNA-sequencing (ddRAD-seq) on a large cohort of *R. irregularis* SSSLs, including homokaryons and dikaryons, to investigate the presence of single nucleotide polymorphism (SNPs), potentially causing qualitative genetic variation among siblings (changes in their DNA sequences compared to the parent). Then, we analysed additional ddRAD-seq data of a small set of dikaryons SSSLs, on which we also performed RNA-sequencing, all to investigate the extent and effects of variation in nucleus genotype ratios.

Interestingly, homokaryon SSSLs also displayed significant differential effects on plant growth when inoculated to cassava plants in field trials conducted in Colombia (Ceballos et al., 2019). Because homokaryon SSSLs should be all genetically identical and could not inherit different nucleus genotypes ratios, these findings suggested the existence of additional sources of trait variation among AMF sibling lines, such as a potentially different epigenetic regulation of the genome. Based on the results from **chapters 2** and **3**, we hypothesized that trait variation among siblings would arise either due to variation in DNA methylation among homokaryon SSSLs, or mostly because of quantitative genetic variation among dikaryon SSSLs. In **chapter 4**, we investigated the potential contribution to trait variation of each of these two factors: (1) DNA methylation and (2) variation in nucleus genotype ratios. Because trait variation among SSSLs in the symbiotic phase, after the fungus and the plant associate, has been more extensively studied and it is affected by a wider range of covariates, we set

up an *in vitro* experiment to test whether *R. irregularis* SSSLs differed in their phenotypic traits even before the AM symbiosis is established, during the process of spore germination and in response to an important environmental cue. We investigated trait variation in response to the environment among 4 homokaryon and 4 dikaryon SSSLs by assessing their percentages of spore germination, the metabolomic profiles of their exudates and their transcriptomes.

References

Angelard C, Colard A, Niculita-Hirzel H, et al. (2010) Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Curr Biol* 20(13): 1216-1221.

Angelard C and Sanders IR (2011) Effect of segregation and genetic exchange on arbuscular mycorrhizal fungi in colonization of roots. *New Phytol* 189(3): 652-657.

Berbee ML and Taylor JW (1993) Dating the evolutionary radiations of the true fungi. *Canadian Journal of Botany* 71(8): 1114-1127.

Bewick AJ, Hofmeister BT, Powers RA, et al. (2019) Diversity of cytosine methylation across the fungal tree of life. *Nature Ecology & Evolution* 3(3): 479-490.

Biermann B and Linderman RG (1983) Use of Vesicular-Arbuscular Mycorrhizal Roots, Intraradical Vesicles and Extraradical Vesicles as Inoculum. *The New Phytologist* 95(1): 97-105.

Bitterlich M, Rouphael Y, Graefe J, et al. (2018) Arbuscular Mycorrhizas: A Promising Component of Plant Production Systems Provided Favorable Conditions for Their Growth. *Front Plant Sci* 9: 1329.

Bronstein JL (1994) Conditional outcomes in mutualistic interactions. *Trends Ecol Evol* 9(6): 214-217.

Brundrett MC (2004) Diversity and classification of mycorrhizal associations. Article in Biological Reviews.

Brundrett MC and Tedersoo L (2018) Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytol* 220(4): 1108-1115.

Bruns TD, Corradi N, Redecker D, et al. (2018) Glomeromycotina: what is a species and why should we care? *New Phytologist* 220(4): 963-967.

Ceballos I, Mateus ID, Peña R, et al. (2019) Using variation in arbuscular mycorrhizal fungi to drive the productivity of the food security crop cassava. *bioRxiv*. DOI: 10.1101/830547. 830547.

Chagnon P-L (2014) Ecological and evolutionary implications of hyphal anastomosis in arbuscular mycorrhizal fungi. *FEMS Microbiology Ecology* 88(3): 437-444.

Chen ECH, Morin E, Beaudet D, et al. (2018) High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. *New Phytol* 220(4): 1161-1171.

Corradi N and Brachmann A (2017) Fungal Mating in the Most Widespread Plant Symbionts? *Trends Plant Sci* 22(2): 175-183.

Corradi N and Lildhar L (2012) Meiotic genes in the arbuscular mycorrhizal fungi. *Communicative & Integrative Biology* 5(2): 187-189.

Croll D, Giovannetti M, Koch AM, et al. (2009) Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus Glomus intraradices. *New Phytologist* 181(4): 924-937.

Croll D and Sanders IR (2009) Recombination in Glomus intraradices, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC Evolutionary Biology* 9(1): 13.

Croll D, Wille L, Gamper HA, 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. *New Phytologist* 178(3): 672-687.

Davison J, Moora M, Öpik M, et al. (2015) Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* 349(6251): 970-973.

Des Roches S, Post DM, Turley NE, et al. (2018) The ecological importance of intraspecific variation. *Nature Ecology & Evolution* 2(1): 57-64.

Drew GC, Stevens EJ and King KC (2021) Microbial evolution and transitions along the parasite–mutualist continuum. *Nature Reviews Microbiology* 19(10): 623-638.

Ehinger MO, Croll D, Koch AM, et al. (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(3): 853-861.

Ewald PW (1987) Transmission modes and evolution of the parasitism-mutualism continuum. *Ann N Y Acad Sci* 503: 295-306.

FAO (2017) The future of food and agriculture – Trends and challenges.

Frank AB and Trappe JM (2005) On the nutritional dependence of certain trees on root symbiosis with belowground fungi (an English translation of A.B. Frank's classic paper of 1885). *Mycorrhiza* 15(4): 267-275.

Friese CF and Allen MF (1991) The Spread of Va Mycorrhizal Fungal Hyphae in the Soil: Inoculum Types and External Hyphal Architecture. *Mycologia* 83(4): 409-418.

Gerdemann JW and Trappe JM (1974) The endogonaceae in the Pacific Northwest.

Gianinazzi S, Gollotte A, Binet MN, et al. (2010) Agroecology: the key role of arbuscular mycorrhizas in ecosystem services. *Mycorrhiza* 20(8): 519-530.

Giovannetti M, Azzolini D and Citernesi AS (1999) Anastomosis Formation and Nuclear and Protoplasmic Exchange in Arbuscular Mycorrhizal Fungi. *Applied and Environmental Microbiology* 65(12): 5571-5575.

Giovannetti M, Sbrana C, Strani P, et al. (2003) Genetic Diversity of Isolates of Glomus mosseae from Different Geographic Areas Detected by Vegetative Compatibility Testing and Biochemical and Molecular Analysis. *Applied and Environmental Microbiology* 69(1): 616-624.

Halary S, Malik S-B, Lildhar L, et al. (2011) Conserved Meiotic Machinery in Glomus spp., a Putatively Ancient Asexual Fungal Lineage. *Genome Biology and Evolution* 3: 950-958.

Hart MM and Reader RJ (2002) Host plant benefit from association with arbuscular mycorrhizal fungi: variation due to differences in size of mycelium. *Biology and Fertility of Soils* 36(5): 357-366.

Jabaji-Hare S, Deschene A and Kendrick B (1984) Lipid Content and Composition of Vesicles of a Vesicular-Arbuscular Mycorrhizal Fungus. *Mycologia* 76(6): 1024-1030.

Kiers ET, Duhamel M, Beesetty Y, et al. (2011) Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* 333(6044): 880-882.

Koch AM, Croll D and Sanders IR (2006) Genetic variability in a population of arbuscular mycorrhizal fungi causes variation in plant growth. *Ecology Letters* 9(2): 103-110.

Koch AM, Kuhn G, Fontanillas P, et al. (2004) High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. *Proc Natl Acad Sci U S A* 101(8): 2369-2374.

Kuhn G, Hijri M and Sanders IR (2001) Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature* 414(6865): 745-748.

Luginbuehl LH and Oldroyd GED (2017) Understanding the Arbuscule at the Heart of Endomycorrhizal Symbioses in Plants. *Curr Biol* 27(17): R952-R963.

Marleau J, Dalpe Y, St-Arnaud M, et al. (2011) Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. *BMC Evol Biol* 11: 51.

Masclaux FG, Wyss T, Mateus-Gonzalez ID, et al. (2018) Variation in allele frequencies at the bg112 locus reveals unequal inheritance of nuclei in a dikaryotic isolate of the fungus Rhizophagus irregularis. *Mycorrhiza* 28(4): 369-377.

Mondo SJ, Dannebaum RO, Kuo RC, et al. (2017) Widespread adenine N6-methylation of active genes in fungi. *Nature Genetics* 49(6): 964-968.

Mosse B and Bowen G (1968) A key to the recognition of some Endogone spore types. *Transactions of the British Mycological Society* 51(3-4): 469-483.

Munkvold L, Kjøller R, Vestberg M, et al. (2004) High functional diversity within species of arbuscular mycorrhizal fungi. *New Phytologist* 164(2): 357-364.

Pawlowska TE and Taylor JW (2004) Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature* 427(6976): 733-737.

Peña R, Robbins C, Corella JC, et al. (2020) Genetically Different Isolates of the Arbuscular Mycorrhizal Fungus Rhizophagus irregularis Induce Differential Responses to Stress in Cassava. *Frontiers in Plant Science* 11.

Pirozynski KA and Malloch DW (1975) The origin of land plants: a matter of mycotrophism. *Biosystems* 6(3): 153-164.

Ravnskov S and Jakobsen I (1995) Functional compatibility in arbuscular mycorrhizas measured as hyphal P transport to the plant. *New Phytologist* 129(4): 611-618.

Redecker D, Kodner R and Graham LE (2000) Glomalean fungi from the Ordovician. *Science* 289(5486): 1920-1921.

Remy W, Taylor TN, Hass H, et al. (1994) Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proceedings of the National Academy of Sciences* 91(25): 11841-11843.

Rodriguez A and Sanders IR (2015) The role of community and population ecology in applying mycorrhizal fungi for improved food security. *ISME J* 9(5): 1053-1061.

Rooney DC, Killham K, Bending GD, et al. (2009) Mycorrhizas and biomass crops: opportunities for future sustainable development. *Trends Plant Sci* 14(10): 542-549.

Ropars J, Toro KS, Noel J, et al. (2016) Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nature Microbiology* 1(6): 16033.

Sagan L (1967) On the origin of mitosing cells. *J Theor Biol* 14(3): 255-274.

Sanders I, Alt M, Groope K, et al. (1995) Identification of ribosomal DNA polymorphisms among and within spores of the Glomales: application to studies on the genetic diversity of arbuscular mycorrhizal fungal communities. *New Phytologist* 130(3): 419-427.

Sanders IR (1999) No sex please, we're fungi. *Nature* 399(6738): 737-738.

Sanders IR (2018) Sex, plasticity, and biologically significant variation in one Glomeromycotina species. *The New Phytologist* 220(4): 968-970.

Sanders IR and Fitter AH (1992) Evidence for differential responses between host-fungus combinations of vesicular-arbuscular mycorrhizas from a grassland. *Mycological Research* 96(6): 415-419.

Savary R, Masclaux FG, Wyss T, et al. (2018) A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus Rhizophagus irregularis. *The ISME Journal* 12(1): 17-30.

Schüβler A, Schwarzott D and Walker C (2001) A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research* 105(12): 1413-1421.

Smith SE and Smith FA (2012) Fresh perspectives on the roles of arbuscular mycorrhizal fungi in plant nutrition and growth. *Mycologia* 104(1): 1-13.

Strullu-Derrien C, Selosse M-A, Kenrick P, et al. (2018) The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New Phytologist* 220(4): 1012-1030.

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* 110(50): 20117-20122.

van der Heijden MGA, Klironomos JN, Ursic M, et al. (1998) Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* 396(6706): 69-72.

van der Heijden MGA, Martin FM, Selosse MA, et al. (2015) Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol* 205(4): 1406-1423.

VanValen L (1973) A new evolutionary law. *Evolutionary Theory* 1: 1-30.

Viera A and Glenn MG (1990) DNA Content of Vesicular-Arbuscular Mycorrhizal Fungal Spores. *Mycologia* 82(2): 263-267.

Wipf D, Krajinski F, van Tuinen D, et al. (2019) Trading on the arbuscular mycorrhiza market: from arbuscules to common mycorrhizal networks. *New Phytol* 223(3): 1127-1142.

Wyss T, Masclaux FG, Rosikiewicz P, et al. (2016) Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus Rhizophagus irregularis. *ISME J* 10(10): 2514-2526.

Chapter 2

The methylome of the model arbuscular mycorrhizal fungus, *Rhizophagus irregularis*, shares characteristics with early diverging fungi and Dikarya

Anurag Chaturvedi[1,#] ¶, Joaquim Cruz Corella[1,#], Chanz Robbins[1,#], Anita Loha[2], Laure Menin[3], Natalia Gasilova[3], Frédéric G. Masclaux[1], Soon-Jae Lee[1], Ian R. Sanders[1,*]

[1] Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015, Lausanne, Switzerland. [2] Department of Plant Molecular Biology, University of Lausanne, 1015 Lausanne, Switzerland. [3] Institute of Chemical Sciences and Engineering, Swiss Federal Institute of Technology Lausanne (EPFL), SSMI, Batochime, CH-1015 Lausanne, Switzerland. ¶Present address: Environmental Genomics Group, School of Biosciences, the University of Birmingham, Birmingham, B15 2TT, UK

[*]Corresponding author: Ian R. Sanders

[#]AC, JCC, CR contributed equally to this work

This chapter was published in *Communications Biology* in July 22nd, 2021.

Abstract

Early-diverging fungi (EDF) are distinct from Dikarya and other eukaryotes, exhibiting high N6-methyldeoxyadenine (6mA) contents, rather than 5-methylcytosine (5mC). As plants transitioned to land the EDF sub-phylum, arbuscular mycorrhizal fungi (AMF; Glomeromycotina) evolved a symbiotic lifestyle with 80% of plant species worldwide. Here we show that these fungi exhibit 5mC and 6mA methylation characteristics that jointly set them apart from other fungi. The model AMF, *R. irregularis,* evolved very high levels of 5mC and greatly reduced levels of 6mA. However, unlike the Dikarya, 6mA in AMF occurs at symmetrical ApT motifs in genes and is associated with their transcription. 6mA is heterogeneously distributed among nuclei in these coenocytic fungi suggesting functional differences among nuclei. While far fewer genes are regulated by 6mA in the AMF genome than in EDF, most strikingly, 6mA methylation has been specifically retained in genes implicated in components of phosphate regulation; the quintessential hallmark defining this globally important symbiosis.

Keywords: Symbiosis, Epigenomics, Arbuscular mycorrhizal fungi, N6-methyldeoxyadenine, 5-methylcytosine

Introduction

The early diverging fungi (EDF), comprising several phyla, are ancient and thought to be up to one billion years old (Mondo et al., 2017; Lucking et al., 2009). Their methylation patterns are mostly distinct from the two later-diverging fungal phyla making up the Dikarya, as well as most other eukaryotes, as they exhibit high levels of N6-methyldeoxyadenine (6mA) and very low levels of 5-methylcytosine (5mC) (Mondo et al., 2017). Thus, this major genomic divergence between these two groups during the evolution of the fungal kingdom is also mirrored in their epigenome. Like in bacteria, 6mA in EDF is primarily located at symmetrical ApT motifs in gene bodies; a feature that is associated with active gene transcription and also likely allows inheritance of 6mA epigenetic marks during DNA replication(Mondo et al., 2017). The Dikarya have lost these features of 6mA methylation and, like in many other Eukaryotes, 5mC is considered to play a more important role in gene regulation. The Glomeromycotina are considered as a clade of the EDF (Spatafora et al., 2016). Known as arbuscular mycorrhizal fungi (AMF), what sets them apart from other fungal groups is that they have been entirely symbiotic with terrestrial plants ever since plants colonized land approximately 450 million years ago and are thought to have played a role in this major evolutionary transition(Remy et al., 1994). The most comprehensive studies of fungal 6mA and 5mC methylation have detailed the prevalence of these epigenetic marks across the fungal phylogeny, including both EDF and Dikarya (Bewick et al., 2019; Mondo et al., 2017). Despite being one of the most important fungal clades, no Glomeromycotina species was included in those studies (Fig. 1). While phylogenetic studies clearly place AMF in the Mucoromycota clade of EDF (Bonfante and Venice, 2020), they share some lifestyle and genomic features with the Dikarya. First, they form dikaryons (Ropars et al., 2016). Second, a homeodomain-containing mating locus resembles those found in the Basidiomycota, rather than in the Mucoromycota (Chen et al., 2018a; Ropars et al., 2016). AMF are major drivers of plant diversity and global carbon and nutrient cycles (van der Heijden et al., 2015; Steidinger et al., 2019; van der Heijden et al., 1998). Numerous studies have shown the capacity of AMF to increase plant P and N acquisition as well as growth(Smith and Read, 2008; Hodge and Fitter, 2010). Inoculation of globally important crops with AMF increases overall yields (Zhang et al., 2019), while reducing the necessity for P fertilization (Ceballos et al., 2019). Given the rapidly depleting global stocks of P (Gilbert, 2009; Gross, 2017), this highlights their importance for global food security. While studies have revealed important characteristics of AMF genomes, little is known to date on their methylome nor are there any insights into its functional implications.

The model AMF species *Rhizophagus irregularis* has been the focus of genome sequencing studies (Tisserant et al., 2013). Large intra-specific genome differences exist within this fungus (Koch et al., 2004; Ropars et al., 2016; Wyss et al., 2016; Chen et al., 2018b; Savary et al., 2018). This variation is considered biologically important as it leads to significant differences in fungal quantitative growth traits, P uptake and transfer to plants, and plant productivity (Koch et al., 2004; Munkvold et al., 2004; Koch et al., 2006; Ceballos et al., 2019). However, some *R. irregularis* isolates are coenocytic

homokaryons, harbouring multiple identical haploid nuclei, while others are coenocytic dikaryons, harbouring a population of two haploid nucleus genotypes (Ropars et al., 2016; Chen et al., 2018a; Masclaux et al., 2019). Clonal offspring of *R. irregularis* dikaryons have been shown to enormously alter the growth of rice by up to five-fold and cassava by up to three-fold (Angelard et al., 2010; Ceballos et al., 2019). Such variation could be due to inheritance of different proportions of the two nucleus genotypes (Angelard et al., 2010; Masclaux et al., 2018). However, clonal offspring of homokaryon *R. irregularis*, that are genetically identical, also induce equally large differences in cassava growth; pointing strongly to the role of epigenetic variation in symbiotic effects of these fungi (Ceballos et al., 2019).

Here we used the Pacific Biosciences RSII platform with SMRT cell technology to compare structural variation (SV) and 6mA epigenomic variation among genetically different *R. irregularis* isolates, between two clones and between two nucleus genotypes within a dikaryon isolate. The epigenomic characteristics across the fungal kingdom have previously comprised only one isolate of a species as representative for each of the main fungal clades, probably owing to the high costs of resolving 6mA epigenomes with this platform. However, because of the biological interest in elucidating 6mA methylation differences among isolates, among clones and among nuclei within the fungi, this study requires a particularly large number of SMRT-cells. We performed whole genome sequencing on six *R. irregularis* isolates (A1, A5, B12, C2, C5 and C3). We present a study on the 6mA methylome in AMF, characterising these epigenetic marks in *R. irregularis* isolates C2, C5 and C3. These isolates show strongly differential effects on plant growth (Ceballos et al., 2019), even though C2 and C5 are genetically indistinguishable at the SNP level (Wyss et al., 2016). Isolates C3 and A5 are dikaryons (Masclaux et al., 2018; Ropars et al., 2016; Masclaux et al., 2019). Because these fungi have haploid nuclei (Ropars et al., 2016), single-molecule sequencing revealing variation within a dikaryon represents variation between the two nucleus genotypes.

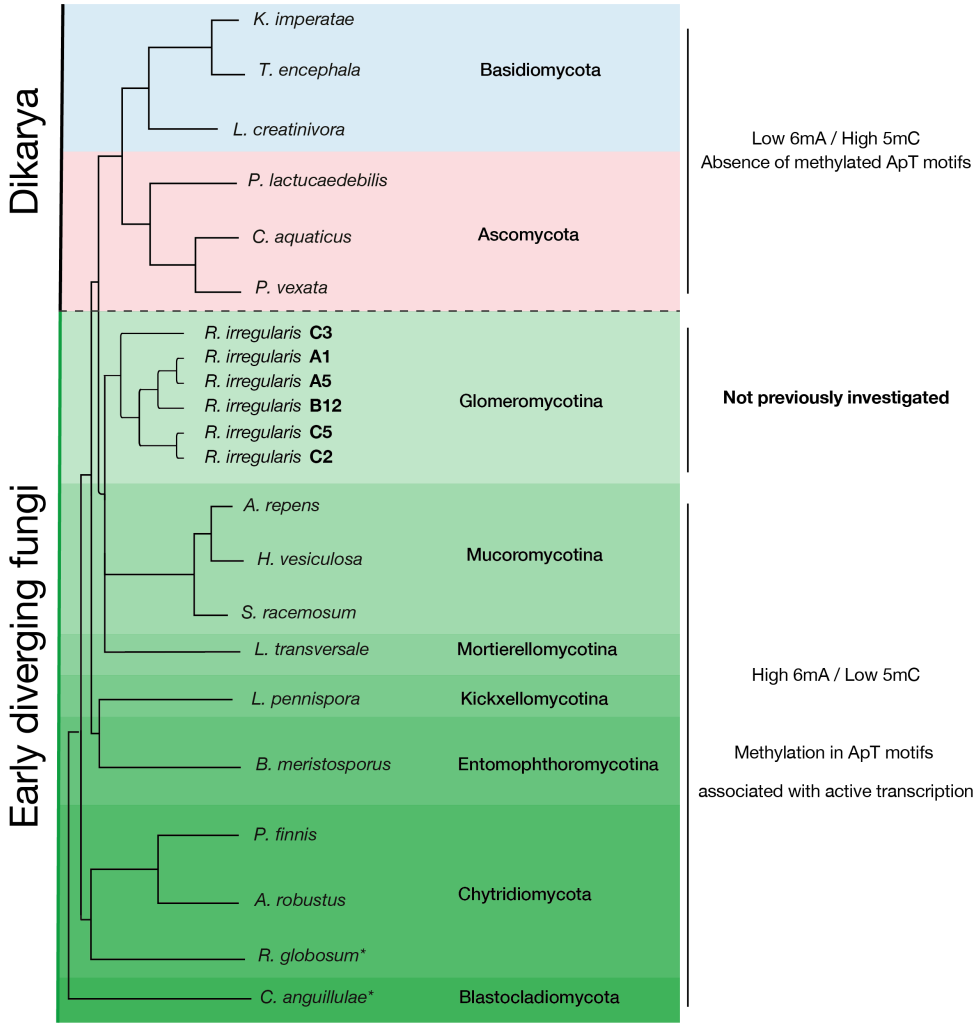


Figure 1. A phylogeny of the fungal kingdom showing that at the major divergence of fungi between Dikarya from early diverging fungi (EDF) there was a shift in the functional methylome from high 6mA and low 5mC contents to high 5mC and low 6mA contents. However, the arbuscular mycorrhizal fungi (Glomeromycotina) have not been previously investigated. *Represents species that do not follow the general pattern of methylation in EDF. In these species, 6mA is low and appears not to have the same functional significance as in EDF as it does not occur in ApT motifs, which are important for ensuring active gene transcription.

Materials and Methods

DNA isolation, library preparation and sequencing

A mixture of fungal hyphae and spores from each of the *R. irregularis* isolates was used for all DNA extractions. The biological material was produced *in vitro* and either provided by Symbiom (Lanškroun, Czech Republic) (A5, B12, C2, C3, C5) or produced in-house using a split-plate culture system (A1) (Rosikiewicz et al., 2017). Material of each of the isolates was collected at the same time and stored at 4°C in sterile ddH_2O until use.

In a laminar flow hood, we took 1 mL of a spore suspension and as much water was removed as possible, before being flash frozen. Each sample was pulverized using cryogenic grinding (precooling: 2 min. 5 Hz; grinding1: 30 sec. 25 Hz; intermission: 30 sec. 5 Hz; grinding 2: 30 sec. 25 Hz) with a Retsch CryoMill (9.739 299, Retsch GmbH, Germany) and were stored in liquid Nitrogen in preparation for DNA extraction.

DNA extraction was performed using the MagAttract® HMW DNA Mini Kit (67563 Qiagen) in exactly the same way for each isolate. Briefly, the protocol was scaled up to account for starting material of around 100 mg. After MB buffer was added, the protocol was modified to increase incubation time at 25°C to 10 min. at 800 rpm. After the removal of the first supernatant, volumes were used as written in the protocol, and only incubation times were increased to 2 min. at 800 rpm. The last two steps of the protocol were performed according to the manual and cut tips were used for transferring high molecular weight DNA. Individual extractions were then pooled for each individual and purified once with Agencourt AMPure XP® (Beckman Coulter) magnetic beads using 1X volumes of beads. DNA was then quantified with Quantus™ Flurometer (Promega) and analyzed for fragment length using Fragment Analyzer (Advanced Analytical Technologies, Inc.).

High molecular weight DNAs were sheared in a Covaris g-TUBE (Covaris, Woburn, MA, USA) to obtain approximately 20 kb fragments. After shearing the DNA size distribution was checked on a Fragment Analyzer (Advanced Analytical Technologies, Ames, IA, USA). 5 µg of the sheared DNA was used to prepare a SMRTbell library with the PacBio SMRTbell Template Prep Kit 1 (Pacific Biosciences, Menlo Park, CA, USA) according to the manufacturer's recommendations. The resulting libraries were size selected on a BluePippin system (Sage Science, Inc. Beverly, MA, USA) for molecules larger than 15 kb. The resulting libraries were sequenced with P4/C2 chemistry and MagBeads on a PacBio RSII system (Pacific Biosciences, Menlo Park, CA, USA) at 240 min movie length using SMRT cells v2.

Genome assemblies and annotation

The genome assemblies of homokaryon isolates (A1, B12, C2, C5) and dikaryon isolates (A5 and C3) were conducted using two separate bioinformatics software i.e. HGAP4 and FALCON-Unzip, respectively. Specifically, HGAP4-based assemblies of A1, B12, C2 and C5 (Chin et al., 2013) were scaffolded using SSPACE (Boetzer and Pirovano, 2014) followed by polishing with PILON (Walker et al., 2014) using Illumina reads from (Chen et al., 2018b). BlobTools (Laetsch and Blaxter, 2017) was used for the identification and removal of non-fungal sequence contaminations. The resulting scaffolds were renamed in descending order based on their length. Sequence data of the dikaryon isolates A5 and C3 was first assembled using FALCON-Unzip (Chin et al., 2016), which is a diploid-aware genome assembler that generates two genome assemblies consisting of one primary genome and a secondary haplotig. This software uses SMRT long-read data to identify heterozygous regions in the genome and constructs haplotigs. Thereafter, the resulting assemblies were similarly scaffolded as homokaryons using SSPACE and polished using PILON. After removal of non-fungal sequence contaminations using BlobTools, scaffolds were sorted and renamed according to their length. Augustus (Stanke et al., 2008) was trained with publicly available RNA sequencing data of C2, C3 and DAOM197198 (Masclaux et al., 2019) and then used for gene predictions on all assemblies. The functional GO term categories were identified using reciprocal BLAST (Camacho et al., 2009) searches for genes in the published gene annotations of AMF isolates (Chen et al., 2018b). The resulting assemblies and annotations are accessible through the European Nucleotide Archive under the accession number: PRJEB33553.

The phylogeny of six *R. irregularis* isolates was constructed as following. We first identified conserved gene families with one gene per isolate using OrthoFinder(Emms and Kelly, 2019) (n=6941). Each single copy gene from each family were then aligned with MAFFT software using default settings (Katoh and Standley, 2013). Further cleaning and concatenation of alignments were performed using GBLOCKS v0.91b (Talavera and Castresana, 2007). RAxML (Stamatakis, 2014) was then used to calculate phylogenetic distances using the PROTGAMMAWAG protein model and trees were constructed by bootstrapping with 1000 iterations to provide node support and visualized in dendroscope (Huson et al., 2007) and geneious (Kearse et al., 2012).

Mapping of raw reads and variant calling for structural variant detection

To detect and report the structural variation (defined in this study as > 30 bp) among the six isolates of *R. irregularis* (A1, A5, B12, C2, C5 and C3) we choose to compare structural variation to a reference isolate DAOM197198 (Chen et al., 2018b). While the threshold length for detection of an SV is arbitrary, we used a threshold of > 30 bp as this is in line with contemporary studies using long-range sequence data(Spielmann et al., 2018; Alonge et al., 2020). We used CoNvex Gap-cost alignMents for Long Reads (NGMLR) to align the raw reads on the reference genome, and Sniffles to call the structural variation from each of the alignment files (Sedlazeck et al., 2018). We ran both tools

with default parameters and filtered for translocations and all variants that had a low read support (RE<15).

In addition, higher sequencing depth in isolate C3 allowed the analysis of SV within a dikaryon isolate of AMF. In order to identify these SVs, we first mapped raw sequencing data of C3 on the primary assembly of C3 using NGMLR. Variants were then called by Sniffles and the same filtering described above was applied to discard translocations and variants with low read support.

Characterization of structural variation in *R. irregularis*

First, we counted the number of structural variations that were reported in the callset of each isolate for the four following classes: insertions (INS), deletions (DEL), inversions (INV) and duplications (DUP).

Second, we computed the Jaccard distances among the isolates by performing all possible pair-wise comparisons of their SV callsets. We accounted for the number of SVs with a reciprocal overlap of at least 80% (BEDtools) (Quinlan, 2014) and of the same SV class. Hierarchical clustering was then performed based on the Jaccard distances with the R package "pvclust" (Suzuki and Shimodaira, 2006), by bootstrapping with 10000 iterations.

Third, in order to find out which structural variants were related to the presence of repeated sequences in the genome, we used the available repeat annotation of isolate DAOM197198 using the BEDTools suite. TE-related SVs were determined by the presence of their breakpoints within the span of an annotated TE in isolate DAOM197198. Finally, gene annotation of isolate DAOM197198 was used to determine the genes intersecting SVs.

In the case of the SV within the dikaryon isolate C3, we computed the same metrics described above. Variants related to repeated sequences were determined on the basis of repeat annotation of C3 primary assembly. Similarly, genes directly affected by SVs were detected on the basis of the gene annotation model of C3.

Comparison of SVs between isolates C2 and C5

Because previous studies had documented the high genetic relatedness of the two *R. irregularis* isolates C2 and C5, we investigated more in detail whether or not SVs existed between them to best determine whether they were actually clones. Two different *in silico* approaches were used to investigate possible structural differences between the genomes of these two isolates.

First, we compared the two final SV callsets of each isolate generated after comparison with the reference isolate DAOM197198. We counted the number of SVs that were co-occurring in the same locations in both isolates. Then, we manually curated all the SVs that initially appeared to be isolate specific using IGV.

Second, in order to reduce the chance of observing false positives, we repeated the mapping of the raw sequence data of C2 and C5 using NGMLR to the newly assembled genome of C2. Then, SV was

called using Sniffles with default parameters. Finally, we manually curated all the SVs that appeared to be isolate specific using IGV.

Epigenetic modification detection and data analysis

The Pacific biosciences RSII platform reads were converted from native bax format to subread bam format using bax2bam SMRT Analysis software. Next, pbalign software was used for mapping subreads to genome assemblies of C2, C5 and C3 using algorithm options to be "--bestn 10 --minMatch 12 --maxMatch 30 --minSubreadLength 50 --minAlnLength 50 --minPctSimilarity 70 --minPctAccuracy 75 --hitPolicy randombest --concordant --randomSeed 1 --useQuality --minPctIdentity 70.0". Tools analysing polymerase kinetics were then used for the detection of 6mA signatures with minimum sequencing 25x per strand coverage with p-value of 0.001. All the tools can be accessed at: https://github.com/PacificBiosciences.

All detected 6mA signatures were analysed for the degree of variation (methylated AT content divided by total AT content), topology (localization of ApT motifs within the gene body) and functional annotation (based on GO analysis). Counting the number of genes affected by methylation, we only considered genes to be methylated which harbour ApT di-nucleotide motif (with a mean value of 10 ApT di-nucleotide motifs per gene; Supplementary Fig. S7). The BEDTools suite was used for the characterization of 6mA topology along the gene body. The gene body includes the start site of the gene to the end site of the gene with a 500 bp upstream (promoter region) and a 500bp downstream region as used in Mondo et. al(Mondo et al., 2017). There is a lack of complete information regarding gene boundaries that includes regulatory regions in *R. irregularis,* thus, the limit that represents the gene body has to be set arbitrarily.

To test whether 6mA presence was positively correlated with gene transcription in *R. irregularis*, we analysed transcriptome data generated from isolates C2 and C3(Masclaux et al., 2019) and the growth conditions for both RNA-seq and DNAseq (long-read sequencing) were exactly the same. RNA sequencing data was mapped on the genome assemblies of each isolate using Tophat(Trapnell et al., 2009). Cufflinks(Roberts et al., 2011) was used for transcript assembly and transcript abundance analysis, using default parameters and isolate-specific gene models, respectively. Statistical analysis was performed using R version 3.5.3. Kendall's correlation coefficient τ and statistical significance were calculated between 6mA density in the gene (number of sites of 6mA / length of coding region for the gene) and gene expression (FPKM, Fragments Per Kilobase Million) by package "Kendall". Statistical significance was assumed at the 95% level.

LC-MS/MS analysis

Approximately 1 µg genomic DNA was sequentially digested to render individual nucleosides. After incubation at 95°C for 3 min., DNA was cooled on ice for 1 min. and digested overnight at 42°C (1 U DNase I from bovine pancreas with included buffer; Roche Diagnostics, Indianapolis, USA). Following digestion, 3.4 µL of 1M NH_4HCO_3 and 1 U phosphodiesterase (from *Crotalus adamanteus*

venom and reconstituted in: 110 mM Tris HCl (pH 8.9), 110 mM NaCl, 15 mM $MgCl_2$; Sigma-Aldrich) were added to break phosphodiester bonds and incubated at 37°C for 2h. Immediately after, 2 U alkaline phosphatase (from bovine intestinal mucosa; Sigma-Aldrich) was added followed by another round of incubation at 37°C for 2h. Fully digested nucleosides were passed through 0.22 µm cellulose acetate filter (Costar® Corning Inc., Salt Lake City, USA), to remove enzymes. After the first round of filtration, the internal standard N^{15}-labeled form of m4dC was added to samples and then filtered one final time using Ultrafree-MC GV Durapore (Merck Millipore, Cork, Ireland).

Quantitative analysis of global levels of dC, m5dC, dA and m6dA was performed on a high-resolution Q Exactive HF Orbitrap FT-MS instrument (Thermo Fisher Scientific, Bremen, Germany) coupled to a Dionex UltiMate 3000 UPLC system (Thermo Fisher Scientific, Bremen, Germany). Analytes were separated on an Acquity UHPLC HSS T3 column (100 × 2.1 mm, 1.8-µm particle size). The mobile phase consisted of 0.1% aqueous formic acid (solvent A) and 0.1% formic acid in acetonitrile (solvent B) at a flow rate of 300 µl/min. Calibration curves (ranging from 0.1 nM to 500 nM) were generated with serial dilutions of synthetic standards of target compounds. The mass spectrometer was set in positive-ion mode and operated in parallel reaction monitoring. Ions of masses 228.10 (dC), 242.11 (m5dC), 252.11 (dA) and 266.12 (m6dA) were isolated with 2 m/z isolation window and fragmented by HDC (higher-energy collisional dissociation) with NCE (normalized collision energy) of 28%. Full MS/MS scans were acquired with resolution 30000 for the base fragments 112.0508, 126.0664, 136.0617 and 150.0774 m/z (cytosine, methylcytosine, adenine and methyladenine, respectively) with 5 p.p.m. mass. The extracted ion chromatogram of the base fragment was used for quantitation. The accurate mass of the corresponding base fragment was extracted with the XCalibur Qual Browser and XCalibur Quan Browser software (Thermo Scientific) and used for quantification. In addition, peak area integration was manually curated and corrected when necessary. All measurements were performed in duplicate and quantities were represented as mean values of two replicates. To estimate the percentage of methylated cytosine and methylated adenine, we divided the value of methylated cytosine and methylated adenine, quantified in nM by the total amount of cytosine and adenine (methylated and non-methylated) present in the sample, respectively.

Identification of putative DNA 6mA methyltransferases in *R. irregularis*

We investigated all MT-A70 family hits (PF05063, pfam version 33.1) in 22 fungal species from Mondo et al., 2017)(Mondo et al., 2017) and 11 AMF isolates and species (6 isolates from the present study and 5 AMF species for which genomes were available from the joint genome institute (JGI) and Genbank. The sources of 5 AMF species from JGI and Genbank are the following: *Rhizophagus cerebiforme* DAOM 227022 (https://mycocosm.jgi.doe.gov/Rhice1_1), *Rhizophagus diaphanous* (https://mycocosm.jgi.doe.gov/Rhidi1), *Rhizophagus clarus* (GCA_003203555.1), *Gigaspora margarita* (GCA_009809945.1), *Gigaspora rosea* (GCA_003550325.1). The protein family search was conducted by hmmsearch, from HMMR v3.0(Wheeler and Eddy, 2013), with default parameters and retained only significant hits (e-value < 0.05, for the full sequence; e-value < 0.01, for the best 1st

domain). Then, we retrieved the protein sequences from their corresponding fungal catalogue, and performed a multiple sequence alignment using MUSCLE(Edgar, 2004) with default parameters. Protein sequences of, *A. repens* and *A. robustus* AMTs (MT-A70 MTases) described by Wang et. al. (2019)(Wang et al., 2019) were also included in the alignment and *T. thermophila* AMTs were used for internal references to define each sub-class of AMTs. The phylogenetic analysis of these proteins was performed using MEGA X(Kumar et al., 2018), where we used the sequence alignment file to construct a maximum likelihood tree, using the default parameters.

Gene ontology enrichment and analysis of structural variation and 6mA in candidate genes

We performed gene ontology (GO) enrichment analysis to find out whether there was an over-representation of given biological processes (BP) and molecular functions (MF) among the genes that presented structural variation within their sequences and those that contained 6mA epigenetic marks. Published DAOM197198 GO annotations were used for GO enrichment of genes containing SV. Published GO annotations of isolate C2 (representing both isolates C2 and C5 because they are genetically indistinguishable) and published isolate A4 GO annotations (representing isolate C3 because A4 and C3 are genetically indistinguishable) were used as a background for GO analysis of genes containing 6mA epigenetic marks. The BINGO plugin of Cytoscape (Shannon et al., 2003; Maere et al., 2005)was used for this purpose. The GO terms with $P \leq 0.05$, after Benjamini & Hochberg (FDR) correction, were considered to be over-represented Biological Process (BP) and Molecular Function (MF) categories. In addition, we investigated individual genes for presence/absence of SV and 6mA epigenetic modification in a set of P and sugar transporters and other genes that are involved in P acquisition and metabolism. The annotation of these genes was based on molecular functions in GO categories and sequenced based blast analysis using AMF genes (Ezawa and Saito, 2018; Helber et al., 2011; Xie et al., 2016).

Statistics and reproducibility

For LC-MS/MS based quantification of 6mA and 5mC, at least 1 µg of genomic DNA was used for sample preparation for all six *R. irregularis* isolates (A1, A5, B12, C2, C5 and C3). Quantification of 6mA and 5mC was performed in triplicate or duplicate and quantities were represented as mean values of the replicates. The percentage estimation of methylated cytosine and methylated adenine was performed by dividing the value of methylated cytosine and methylated adenine, quantified in nM by the total amount of cytosine and adenine (methylated and non-methylated) present in the sample, respectively.

The Kendall's correlation between gene 6mA methylation density (methylation abundance divided by gene length) and transcription levels was performed using R software while maintaining significance of $P < 0.001$.

Results and Discussion

Genome coverage, annotation and structural variation

Details of genome coverage, annotation, assembly and SV are given in: Supplementary Notes 1 and 2, Supplementary Figs. S1-S5, Supplementary Tables S1-S6 and Supplementary Data 1. Analysis of SV confirmed the previously published phylogenies and very high relatedness between isolates C2 and C5 (Savary et al., 2018; Wyss et al., 2016) (Supplementary Fig. S1; Supplementary Note 3). Analysis of SV also allowed us to show that C2 and C5 are genetically indistinguishable and were, therefore, considered as clones (Supplementary Fig. S4). It also allowed us to define the structural differences between nuclei in the dikaryon C3 (Supplementary Note 4). SV between nucleus genotypes shared the same characteristics as seen between isolates and gene SV was as large as that seen, between some isolates (Supplementary Fig. S5a-d).

The amounts of 6mA and 5mC methylation in arbuscular mycorrhizal fungi are unlike the other EDF

Having confirmed that C2 and C5 were likely clones, and defined the structurally different regions of the two nucleus genotypes in C3, we characterised 6mA methylation. Sufficient coverage (25x per strand) allowed reliable detection of 6mA in C2, C5 and C3 using Pacific Biosciences SMRT Analysis software kineticsTools (Supplementary Fig. S2).

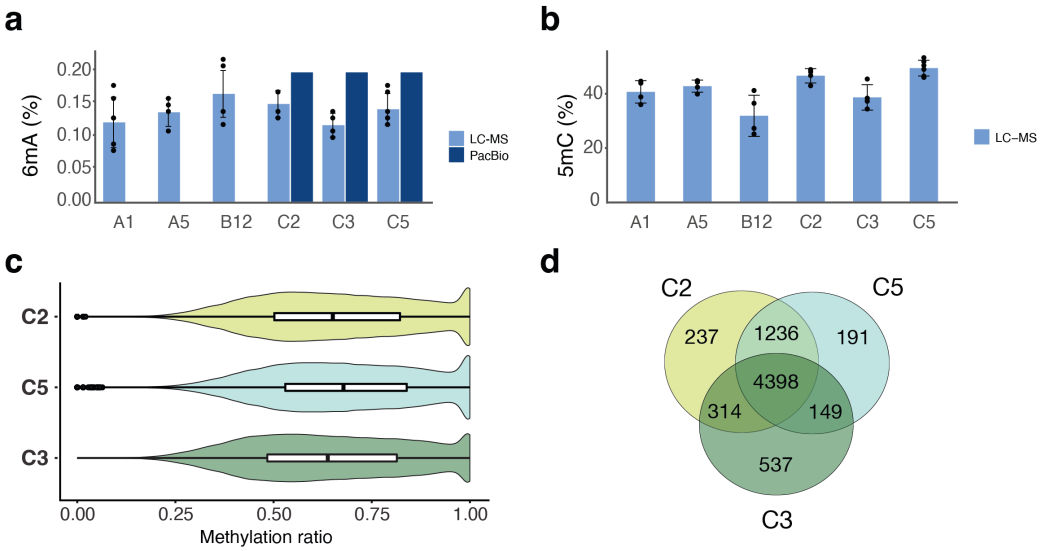


Figure 2. Quantification of 6mA and 5mC in *R. irregularis* isolates (a) Percentage of total methylated adenine (6mA) detected by single molecule DNA sequencing in three isolates (C2, C5 and C3) and validated by LC-MS in six isolates. The data are available in Supplementary data 3. (b) Percent of total methylated cytosine (5mC) detected by LC-MS in six isolates. The data are available in Supplementary data 3. (c) Violin plot (with boxplot inside) showing the distribution of the methylation ratio of *R. irregularis* isolateC2 (n=291654; minima:0, maxima:1, mean:0.6594, Q1:0.5020, Q3:0.8220), C5 (n=240110; minima:0 , maxima:1 , mean:0.6813, Q1:0.5300, Q3:0.8390) and C3 (n=263696; minima:0 , maxima:1 , mean:0.6465, Q1:0.4840, Q3:0.8140) suggesting that 6mA methylation in R. irregularis was highly heterogeneous. (d) Venn diagram representing numbers of common and unique genes harbouring 6mA ApT motifs in isolates C2, C5 and C3. The majority of methylated genes were shared among isolates, yet considerable differences were observed among isolates and between clones C2 and C5.

Approximately 0.2% of adenine was methylated in *R. irregularis* (Fig. 2a). Liquid chromatography–mass spectrometry (LC-MS) confirmed similar proportions of 6mA methylation among isolates (0.12 to 0.17%) (Fig. 2a). Although the Glomeromycotina are considered EDF, 6mA methylation in *R. irregularis* was much lower than most of the rest of this ancient fungal group and was in the range typically seen in Dikarya, plants and animals (Xiao et al., 2018; Fu et al., 2015; Zhang et al., 2015; Greer et al., 2015; Mondo et al., 2017)(Fig. 2a). High abundance of 5mC methylation is expected when 6mA levels are low (Bewick et al., 2019; Mondo et al., 2017) and, indeed, we observed 32.5%-49.5% of cytosine was methylated. This characteristic was consistent with the Dikarya and other Eukaryotes, in the sense that it is high, but not with the other EDF(Mondo et al., 2017) (Fig. 2b). However, the percentage of 5mC exceeds that recorded in other Dikarya to date [5]. There are two species in the EDF where a similarly low 6mA levels have been observed (*Rhizoclosmatium globosum* and *Catenaria anguillula*), but unlike *R. irregularis* both show very low levels of 5mC methylation(Mondo et al., 2017). Higher prevalence of 5mC methylation than in other fungal species suggests an important role of this epigenetic mark in gene regulation in AMF and, thus, requires further characterisation.

Among-nucleus 6mA methylation heterogeneity is a conserved feature

While 6mA methylated sites in most other EDF show full methylation, 6mA methylation in *R. irregularis* was highly heterogeneous, with methylation ratios from 0.64 in C3 to 0.68 in C5 (Fig. 2c). Coverage had a negligible influence on the ratio (Supplementary Fig. S6). Some 6mA heterogeneity has been reported in one member of the EDF (*Lobosporanguim transversal*; Mortierellomycotina), but not to the amount observed in *R. irregularis*. Pairwise comparisons of methylation ratios between isolates showed that only a negligible proportion of methylated sites exhibited a significantly diverging methylation ratio, highlighting the robustness of these methylation heterogeneity estimates. Only 0.62%, 1.41% and 1.29% of methylated sites revealed diverging methylation ratios between C2 and C5, between C2 and C3 and between C5 and C3, respectively (at 2.5% and 97.5% confidence limits). Conducting single-molecule sequencing means that each sequenced DNA strand originates from one haploid nucleus. The haploid status of nuclei in *R. irregularis* means that the observed epigenetic heterogeneity in AMF must be partitioned among the population of nuclei, thus giving a variable function among nuclei within one coenocytic fungus. 6mA methylation heterogeneity among cells in prokaryotes allows population-level adaptation to arise from clonal colonies (Casadesus and Low, 2013; Beaulaurier et al., 2015), and in eukaryotes, among-cell methylation heterogeneity is crucial for differentiation (Heng et al., 2009),(Zhang et al., 2013). Thus, if 6mA has a role in gene regulation, the pattern of among nucleus 6mA heterogeneity represents a layer of variation in *R. irregularis* that could explain the extremely plastic response in the symbiotic effects of clonal sibling lines of these fungi on plant growth and allow a plastic response of the fungus to the environment.

6mA methylation occurs in a core fungal gene set but methylation differences also occur among clones and also between nuclei in dikaryon isolates

Because both genetically different and identical *R. irregularis* isolates show differential effects on plant growth we investigated their 6mA methylation differences. Specifically, 22.8%, 22.8% and 21.9% of genes in C2, C5 and C3 respectively harboured 6mA. Out of a 290 BUSCO core fungal gene set, 270 of those genes occurring in *R. irregularis* were methylated. A total of 7062 genes were methylated in the three isolates (Fig. 2d). All genes methylated in at least one isolate were genes occurring in all isolates. Thus, 6mA methylation acts on a core gene set, similar to other EDF, rather than isolate-specific genes. Most of these (4398 genes; 62%) were methylated in all isolates (Fig. 2d). However, 31% were differentially methylated between C3 and C2/C5. The methylation repertoire was not identical between clones C2 and C5 (Fig. 2d), meaning that epigenetic variation could cause differential gene regulation between genetically identical isolates. SV analysis between two clones will never show two individuals to be completely identical because of sequencing errors, coverage or problems with the reference assemblies. While the very small number of SVs were discounted as artefacts (Supplementary Note 3), we checked to see if differential methylation in those SVs could result in the 6mA methylation differences observed between C2 and C5. Methylation only occurred in 22 SVs and only one of these contained a gene body. Thus, differential methylation in artefactual SVs cannot explain the methylation differences between these two isolates. If 6mA regulates gene expression then this could at least partially explain the extremely large differences in symbiotic effects of clones C2 and C5 on crop growth, as well as among clonal siblings (Ceballos et al., 2019). Differential methylation also occurred between nuclei within the dikaryon C3 (Supplementary Note 5).

6mA methylation in AMF has a likely functional role in gene regulation similar to other EDF

While 6mA and 5mC abundances set AMF apart from most other EDF, several characteristics point to a similar role of 6mA in gene regulation in *R. irregularis* as in other EDF rather than Dikarya. First, 6mA marks in *R. irregularis* occurred as symmetrical methylation at ApT di-nucleotide motifs (73% in C3; 68% in C2 and C5), where AATT was the most abundant (Fig. 3a; Supplementary Data 2). This prevalent epigenomic 6mA signature across the EDF and unicellular algae is distinct from the Dikarya and animals (Fu et al., 2015; Mondo et al., 2017; Xiao et al., 2018). Furthermore, just like in other EDF, and unlike plants and animals (Dominissini et al., 2012; Luo et al., 2014), 6mA sites at these ApT motifs were strongly biased to positions within genes and in the upstream promoter region of the genes (~86% of 6mA-ApT sites)(Fig. 3b). These two non-random patterns of 6mA distribution strongly suggest a functional significance as observed in other EDF. Although two species in other EDF clades (*Rhizoclosmatium globosum* and *Catenaria anguillulae*) also have low 6mA abundance, they either have low, or no, ApT motif methylation (Mondo et al., 2017). That ApT methylation is symmetrical means that methylation can be retained on parental strands during DNA replication and nucleus division and is, thus, likely heritable; a feature that was lost in the later diverging Dikarya.

The non-random patterns of 6mA distribution strongly suggest a likely functional consequence of 6mA symmetrical ApT signatures in the Glomeromycotina epigenome. Indeed, 6mA symmetrical ApT motif signatures (with a mean value of 10 ApT di-nucleotide motifs per gene; Supplementary Fig. S7) in genes were associated with transcription of those genes. Most genes containing 6mA ApT motifs were highly expressed in isolates C2 and C3, irrespective of 6mA abundance (total amount of 6mA) in the gene (FPKM>5) (Fig. 3c), while a significant positive correlation occurred between gene 6mA methylation density (methylation abundance divided by gene length) and transcription levels (Kendall's correlation τ: 0.3649, $P < 0.001$ and τ: 0.4395, $P < 0.001$ in C2 and C3, respectively).

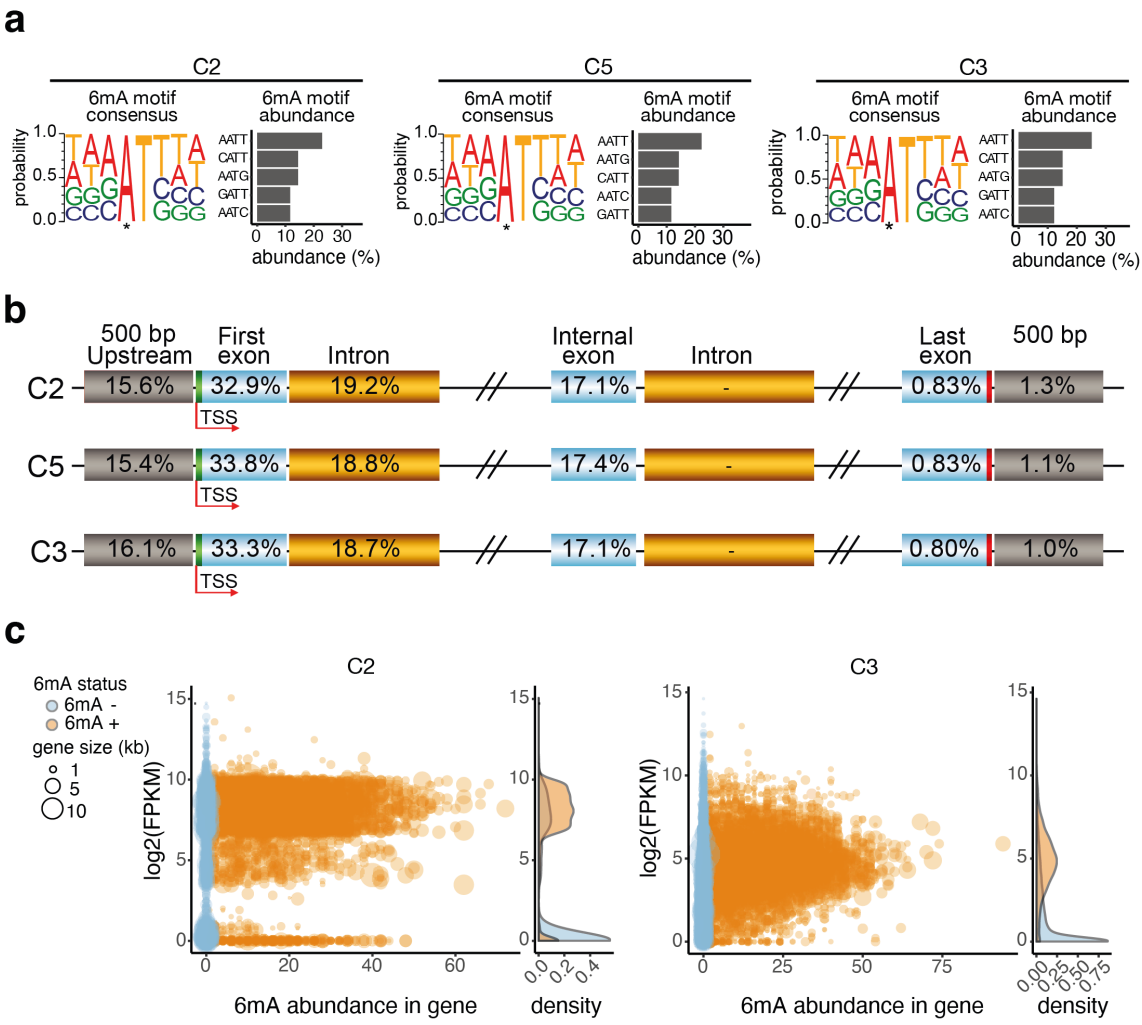


Figure 3. Characterization of 6mA in *R. irregularis* isolates (a) 6mA consensus sequence motifs with the probability of occurrence of each motif and the most abundant motifs in isolates C2, C5 and C3 arranged in decreasing abundance. Several ApT motifs other than AATT, CATT, AATG, GATT and AATC were also present in each of the three isolates and in similar proportions among the three isolates. * Represents the methylated base. Data available in Supplementary Data 2. (b) Topological distribution of 6mA ApT dinucleotides in the gene structure in C2, C5 and C3. (c) Gene expression (log2(FPKM)) in isolates C2 and C3 of genes based on 6mA ApT abundance within gene bodies (6mA+: orange; 6mA-: blue). Gene size is represented by circle size and density curves representing the occurrence of genes with 6mA abundance (6mA+: orange; 6mA-: blue) with respect to gene expression log2(FPKM) values. The data are available in Supplementary Data 4.

Methyltransferases in AMF are more similar to EDF than Dikarya

Mondo et al. (2017)(Mondo et al., 2017) studied the differences in counts between EDF and Dikarya for pfam domains that are thought responsible for fungal 6mA methylation. They found that 44 pfam domains significantly differed ($p < 0.01$). Among them, the MT-A70 family (PF05063) plays a central role as a DNA 6mA methyltransferase (Greer et al., 2015). A subclade of MT-A70, namely AMT1, is responsible for genome-wide DNA methylation levels and the maintenance of symmetric 6mA methylation in ApT motifs in eukaryotes (Wang et al., 2019). Interestingly, Wang et al. (2019)(Wang et al., 2019) showed that two EDF species had an AMT1 homolog, while one species of Dikarya did not. Because symmetric methylation in ApT motifs is a distinct feature of 6mA methylation in EDF compared to the Dikarya, we performed a phylogenetic analysis of the MT-A70 family across the EDF

and the Dikarya with a larger set of fungal taxa compared to Wang et al. (2019)(Wang et al., 2019) including all the fungal species reported in Mondo et al. (2017)(Mondo et al., 2017), previously classified reference sequences from Wang et al. (2019)(Wang et al., 2019) and six additional Glomeromycotina species (including the six isolates of *R. irregularis*). The previously characterized and classified AMTs protein sequences from Wang et al. (2019)(Wang et al., 2019) were used as internal controls to identify the different MT-A70 sub-classes.

As expected, we found that EDF, including AMF, had a larger set of MT-A70 family proteins compared to Dikarya. Moreover, AMT 1 was found in all EDF and AMF that exhibit substantial symmetric ApT methylation, while no homolog was found in Dikarya (Fig. 4). Two EDF, *C. anguillulae* and *R. globosum,* exhibit no, or extremely low, 6mA methylation in ApT motifs, thus, exhibiting a pattern more similar to Dikarya. We show that *R. globosum* harbours a distinctively short AMT1 homolog (Supplementary Fig. S8), and no AMT1 homolog was found in *C. anguillulae*, which is likely why this species has almost no symmetric 6mA methylation.

All AMF genomes we analysed contained a conserved AMT1 homolog as in other EDF, which could explain the high degree of symmetric methylation at ApT motifs despite the fact that 6mA levels are low in *R. irregularis*. The abundance of 6mA methylation and 5mC methylation of fungal genomes is inversely correlated (Mondo et al., 2017), suggesting a relationship between 6mA MTases and 5mC MTases in fungi, with a higher rate of losing than gaining 5mC MTases when an AMT1 homolog is present (Bewick et al., 2019). *R. irregularis* appears to be an exception where conservation of a highly symmetric ApT methylation is coupled with a high degree of 5mC methylation. The AMF clade has all five conserved 5mC MTase types (DNMT1, DNMT2, DNMT5, DIM-2 and RID), while sister clades of the Mucoromycota phylum have lost DNMT5 and RID (Bewick et al., 2019). Together with the divergent 6mA methyltransferases repertoire reported in our study, the conservation of methyltransferases in *Rhizophagus irregularis* is in line with high degree of 5mC methylation found in this fungus and low 6mA level.

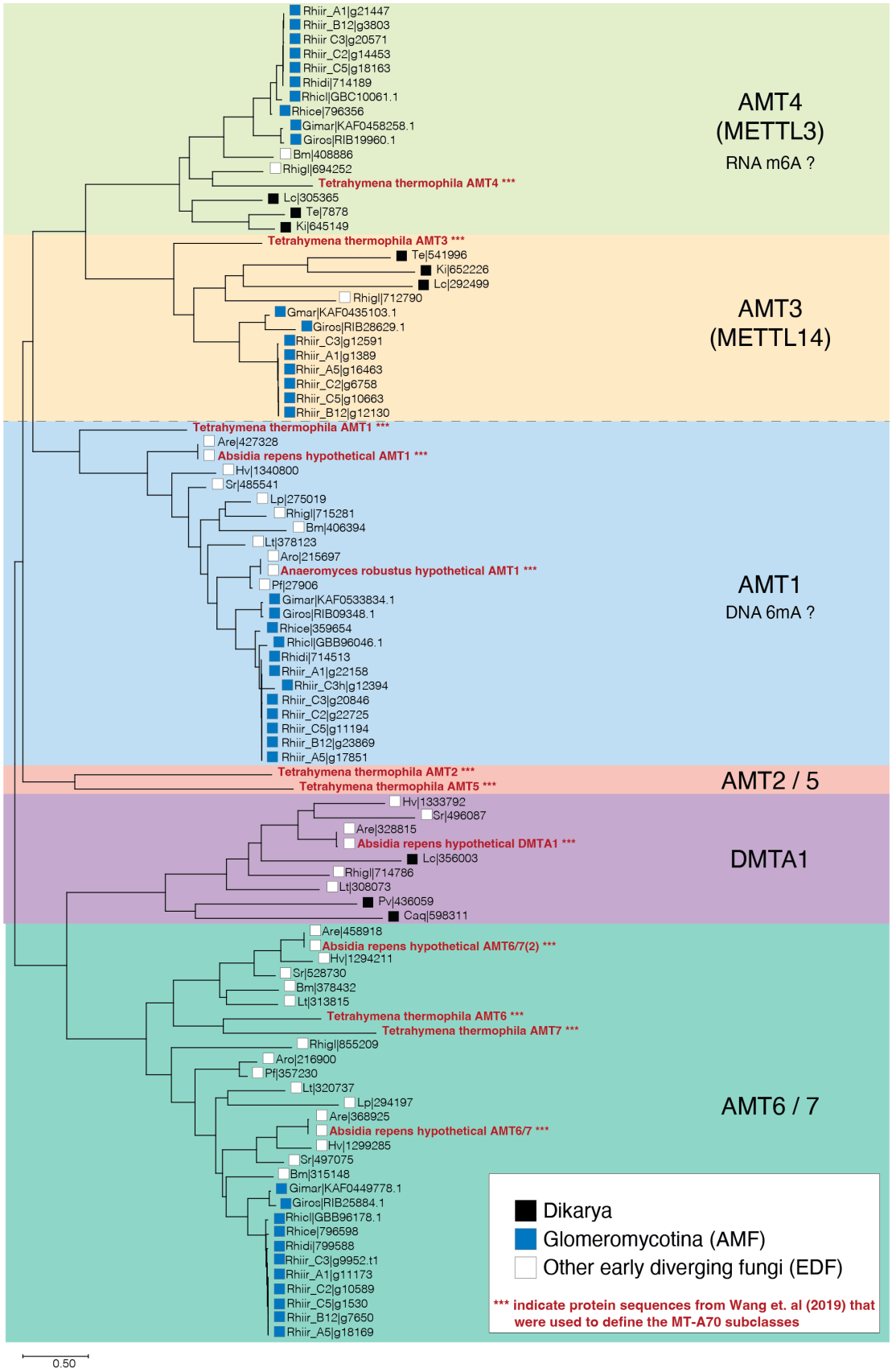


Figure 4. Phylogenetic analysis of MT-A70 proteins across fungal taxa. Putative RNA m6A MTases (METTL3, METTL14) and putative DNA 6mA MTase subclasses (AMT2/5, AMT1, DMTA1 and AMT6/7) are separated by a dashed line. Protein sequences of *T. thermophila*, *A. repens*, and *A. robustus* from Wang et. al. (2019) were used to define the subclasses of MT-A70 proteins. These sequences are coloured in red and marked with three asterisks.

Dikarya, early diverging fungi (EDF) and AMF taxa are specified with black, white and blue squares, respectively. Species abbreviations are as follows: *Clohesyomyces aquaticus* (Caq); *Leucosporidiella creatinivora* (Lc); *Protomyces lactucaedebilis*: (Pl); *Pseudomassariella vexata* (Pv); *Tremella encephala* (Te); *Kockovaella imperatae* (Ki); *Catenaria anguillulae* (Can); *Hesseltinella vesiculosa* (Hv); *Linderina pennispora* (Lp); *Lobosporangium transversale* (Lt); *Piromyces finnis* (Pf); *Syncephalastrum racemosum* (Sr); *Absidia repens* (Are); *Anaeromyces robustus* (Aro); *Basidiobolus meristosporus* (Bm); *Rhizoclosmatium globosum* (Rhigl); *Rhizophagus cerebiforme* (Rhice); *Rhizophagus clarus* (Rhicl); *Rhizophagus diaphanus* (Rhidi); *Rhizophagus irregularis* isolate A1 (Rhiir_A1); *Rhizophagus irregularis* isolate A5 (Rhiir_A5); *Rhizophagus irregularis* isolate B12 (Rhiir_B12); *Rhizophagus irregularis* isolate C2 (Rhiir_C2); *Rhizophagus irregularis* isolate C3 (Rhiir_C3); *Rhizophagus irregularis* isolate C3 (secondary haplotig, pseudo-secondary nucleus genotype) (Rhiir_C3h); *Rhizophagus irregularis* isolate C5 (Rhiir_C5); *Gigaspora margarita* (Gimar); *Gigaspora rosea* (Giros).

6mA gene methylation is biased towards specific gene functions in AMF, including the regulation of DNA methylation

Patterns of among-clone, and within-individual, 6mA heterogeneity could explain why genetically identical AMF clones and siblings of AMF homokaryons display enormous differences in their effects on crop growth. However, while Glomeromycotina 6mA methylation appears to have a similar regulatory role as in other EDF, it affects a much smaller number of genes than in other EDF with notable exception of *Linderina pennispora*. Thus, it was important to investigate whether 6mA methylation affects genes known to affect the functioning of the symbiosis. GO analysis of 6mA harbouring genes showed enrichment of transferase activity, binding, catalytic activity, DNA methyl transferase activity, hydrolase activity (including pyrophosphatase activity) and transporter activity (Fig. 5a; Supplementary Data 1). We also found enrichment of both ion and ATPase transmembrane transporter activity (Fig. 5b and 5c) which is significant given the beneficial effect of AMF is based on the fungus ability to absorb nutrients and transport them to the plant. Additionally, we observed the 6mA methylation of genes enriched for DNA methyltransferase, suggesting that the transfer of methyl groups to DNA (potentially 5mC, as well as 6mA) could itself be under the control of 6mA methylation.

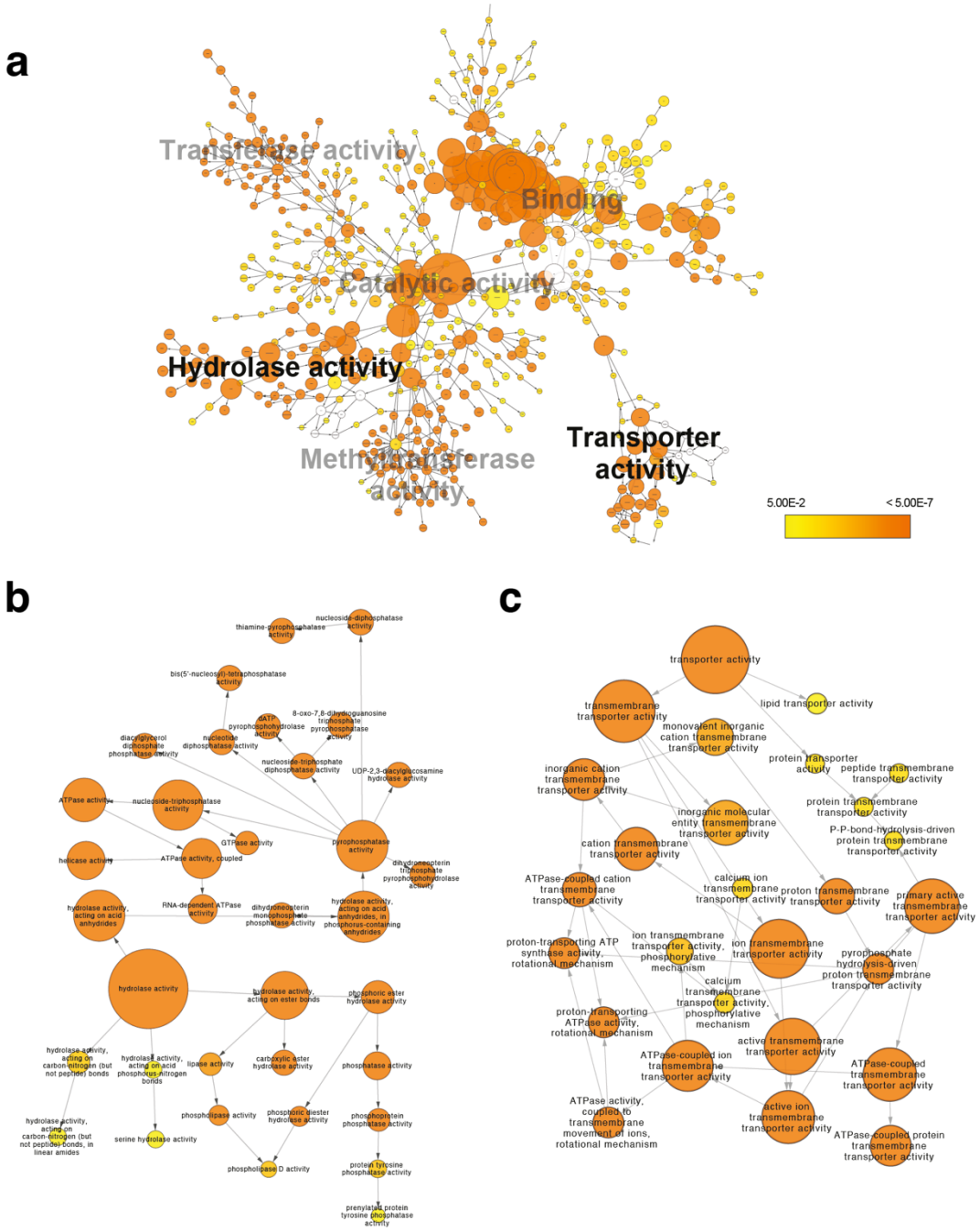


Figure 5. Network of enriched Gene ontology (GO) molecular function categories of commonly methylated 6mA ApT genes in *R. irregularis* isolates C2/C5 and C3. (a) General overview of the six principal enriched clusters of molecular function with the Hydrolase activity and Transporter activity clusters highlighted in bold. (b) Hydrolase activity cluster shown in greater detail. (c) Transporter activity cluster shown in greater detail. The diameter of the circle refers to number of genes per GO term (which is shown in Supplementary Data 1) and the colour of the circle refers to the FDR corrected probability (P), where the yellow to orange gradient represents increasing significance (Supplementary Data 1).

AMF have specifically retained 6mA methylation in genes important for the mycorrhizal symbiosis

The quintessential hallmark of the mycorrhizal symbiosis is that the fungi absorb and transport P to the plant. RNAi and transcription studies have revealed Glomeromycotina genes involved in P transport, metabolism, sensing and signalling, homeostasis and acquisition pathways (Ezawa and Saito, 2018; Xie et al., 2016). Here, we observed that some of these genes contained SV, but the vast majority appear under the regulation of 6mA methylation (Supplementary Table S7; Supplementary Note 6). These included low-affinity SPX domain containing P transporters (PHO87, PHO88, PHO90, PHO91). SPX domains are considered key to P transport and metabolism in eukaryotes including AM fungi (Ezawa and Saito, 2018; Wild et al., 2016). Remarkably, all the genes identified in previous studies as being involved in inositol polyphosphate synthesis/hydrolysis and P responsive signalling appear to be under epigenetic control. Such genes are thought to be extremely important in AMF because these fungi have to take up and regulate P levels much greater than those needed by the fungus alone (Ezawa and Saito, 2018). After P absorption by the fungus, AMF have to convert phosphate into polyphosphate. The vacuolar transporter chaperones VTC1, VTC 2 and VTC4, involved in polyphosphate synthesis (and also containing SPX domains), are also under 6mA epigenetic control. Finally, the parts of the of the PKA signalling pathways, that are strongly upregulated in AMF in response to low P environments, and the MAPK and Tor signalling pathways, that are down-regulated in low P environments appear to be almost completely methylated, likely under epigenetic control on the regulation of these pathways in the fungus in varying P environments. The fact that *R. irregularis* has specifically retained 6mA methylation in the genes regulating the key resource, phosphorus, that the plant needs from this symbiosis, while clearly having lost 6mA methylation from a very large number of other genes is highly intriguing. Partners in mutualistic symbioses are expected to evolve mechanisms to prevent over exploitation by the partner. Given the high 5mC levels in *R. irregularis*, a large part of gene regulation in AMF must be under 5mC control rather than 6mA. Plants have evolved mechanisms to regulate gene expression through 5mC and use mobile small RNA (sRNA) molecules to regulate 5mC methylation (Tamiru et al., 2018). Cross-kingdom transfer of mobile sRNA molecules is a commonly recognized way that organisms in plant – microbe interactions manipulate gene expression of the partner (Huang et al., 2019). Indeed, a recent study of AMF genomes revealed that AMF possess an RNAi system that regulates sRNA, as in plants (Lee et al., 2018). We suggest that if such sRNA transfer occurs between plants and AMF, which is highly likely, then by retaining a 6mA regulatory mechanism to regulate access to the key resource needed by plants, AMF may prevent host manipulation of the genes controlling this resource and allow the fungus to determine the amount of P available to plants.

Conclusions

As Glomeromycotina diverged from other EDF around 500 million years ago and formed symbioses with the first land plants, they evolved epigenomic features akin to the Dikarya, such as high 5mC abundance and low levels of 6mA methylation. However, the combination of 5mC levels higher than those recorded in Dikarya, and low 6mA levels, coupled with 6mA methylation characteristics not seen in other EDF that harbour low 6mA, make the methylome patterns in *R. irregularis* unique in the fungi. Despite the loss of 6mA from a huge number of genes, the Glomeromycotina retained heritable EDF-like 6mA gene regulation capabilities in a conserved core set of genes, including those fundamental to their globally important role in symbiosis and P cycling as well as in those genes controlling methylation of DNA; possibly even 5mC which is the most abundant part of the Glomeromycotina methylome. Methylated adenine in Glomeromycotina is not randomly distributed, occurring symmetrically near genes and is associated with gene transcription. Despite the conservation of 6mA features in the Glomeromycotina, a significant amount of heterogeneity in 6mA methylation among nuclei exists, as well qualitatively differential 6mA among genes in genetically identical clones. These features point to a clear mechanism to explain the enormous differences that genetically indistinguishable AMF and clonal sibling fungi induce in the production of globally important crops.

Data Availability

The raw data, resulting genome assemblies and annotations are accessible through the European Nucleotide Archive under the accession number: PRJEB33553.

Acknowledgements

This work was supported by Swiss National Science Foundation grants (project numbers 31003A_162549/1 and 310030B_182826/1) who's support we gratefully acknowledge.

Author Contributions

AC, FGM, JCC, CR, IRS conceived the study. AC, JCC, CR, SL, FGM performed data analysis. CR, SL, LM, NG performed experimental work. All authors contributed to writing and editing the manuscript.

References

Alonge M, Wang X, Benoit M, et al. (2020) Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell* 182(1): 145-161 e123.

Angelard C, Colard A, Niculita-Hirzel H, et al. (2010) Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Curr Biol* 20(13): 1216-1221.

Beaulaurier J, Zhang XS, Zhu S, et al. (2015) Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. *Nat Commun* 6: 7438.

Bewick AJ, Hofmeister BT, Powers RA, et al. (2019) Diversity of cytosine methylation across the fungal tree of life. *Nat Ecol Evol* 3(3): 479-490.

Boetzer M and Pirovano W (2014) SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* 15: 211.

Bonfante P and Venice F (2020) Mucoromycota: going to the roots of plant-interacting fungi. *Fungal Biology Reviews*. DOI: https://doi.org/10.1016/j.fbr.2019.12.003.

Camacho C, Coulouris G, Avagyan V, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.

Casadesus J and Low DA (2013) Programmed heterogeneity: epigenetic mechanisms in bacteria. *J Biol Chem* 288(20): 13929-13935.

Ceballos IC, Mateus ID, Pena R, et al. (2019) Using variation in arbuscular mycorrhizal fungi to drive the productivity of the food security crop cassava. *bioRxiv*.

Chen EC, Mathieu S, Hoffrichter A, et al. (2018a) Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. *Elife* 7.

Chen ECH, Morin E, Beaudet D, et al. (2018b) High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. *New Phytol* 220(4): 1161-1171.

Chin CS, Alexander DH, Marks P, et al. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10(6): 563-569.

Chin CS, Peluso P, Sedlazeck FJ, et al. (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 13(12): 1050-1054.

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, et al. (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485(7397): 201-206.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5): 1792-1797.

Emms DM and Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20(1): 238.

Ezawa T and Saito K (2018) How do arbuscular mycorrhizal fungi handle phosphate? New insight into fine-tuning of phosphate metabolism. *New Phytol* 220(4): 1116-1121.

Fu Y, Luo GZ, Chen K, et al. (2015) N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. *Cell* 161(4): 879-892.

Gilbert N (2009) Environment: The disappearing nutrient. *Nature* 461(7265): 716-718.

Greer EL, Blanco MA, Gu L, et al. (2015) DNA Methylation on N6-Adenine in C. elegans. *Cell* 161(4): 868-878.

Gross M (2017) Where is all the phosphorus? *Current Biology* 27(21): R1141-R1144.

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. *Plant Cell* 23(10): 3812-3823.

Heng HH, Bremer SW, Stevens JB, et al. (2009) Genetic and epigenetic heterogeneity in cancer: a genome-centric perspective. *J Cell Physiol* 220(3): 538-547.

Hodge A and Fitter AH (2010) Substantial nitrogen acquisition by arbuscular mycorrhizal fungi from organic material has implications for N cycling. *Proc Natl Acad Sci U S A* 107(31): 13754-13759.

Huang CY, Wang H, Hu P, et al. (2019) Small RNAs - Big Players in Plant-Microbe Interactions. *Cell Host Microbe* 26(2): 173-182.

Huson DH, Richter DC, Rausch C, et al. (2007) Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8: 460.

Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4): 772-780.

Kearse M, Moir R, Wilson A, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12): 1647-1649.

Koch AM, Croll D and Sanders IR (2006) Genetic variability in a population of arbuscular mycorrhizal fungi causes variation in plant growth. *Ecol Lett* 9(2): 103-110.

Koch AM, Kuhn G, Fontanillas P, et al. (2004) High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. *Proc Natl Acad Sci U S A* 101(8): 2369-2374.

Kumar S, Stecher G, Li M, et al. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35(6): 1547-1549.

Laetsch D and Blaxter M (2017) BlobTools: Interrogation of genome assemblies. F1000Res.

Lee SJ, Kong M, Harrison P, et al. (2018) Conserved Proteins of the RNA Interference System in the Arbuscular Mycorrhizal Fungus Rhizoglomus irregulare Provide New Insight into the Evolutionary History of Glomeromycota. *Genome Biol Evol* 10(1): 328-343.

Lucking R, Huhndorf S, Pfister DH, et al. (2009) Fungi evolved right on track. *Mycologia* 101(6): 810-822.

Luo GZ, MacQueen A, Zheng G, et al. (2014) Unique features of the m6A methylome in Arabidopsis thaliana. *Nat Commun* 5: 5630.

Maere S, Heymans K and Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21(16): 3448-3449.

Masclaux FG, Wyss T, Mateus-Gonzalez ID, et al. (2018) Variation in allele frequencies at the bg112 locus reveals unequal inheritance of nuclei in a dikaryotic isolate of the fungus Rhizophagus irregularis. *Mycorrhiza* 28(4): 369-377.

Masclaux FG, Wyss T, Pagni M, et al. (2019) Investigating unexplained genetic variation and its expression in the arbuscular mycorrhizal fungus Rhizophagus irregularis: A comparison of whole genome and RAD sequencing data. *PLoS One* 14(12): e0226497.

Mondo SJ, Dannebaum RO, Kuo RC, et al. (2017) Widespread adenine N6-methylation of active genes in fungi. *Nat Genet* 49(6): 964-968.

Munkvold L, Kjoller R, Vestberg M, et al. (2004) High functional diversity within species of arbuscular mycorrhizal fungi. *New Phytologist* 164(2): 357-364.

Quinlan AR (2014) BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* 47: 11 12 11-34.

Remy W, Taylor TN, Hass H, et al. (1994) Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc Natl Acad Sci U S A* 91(25): 11841-11843.

Roberts A, Pimentel H, Trapnell C, et al. (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27(17): 2325-2329.

Ropars J, Toro KS, Noel J, et al. (2016) Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nat Microbiol* 1(6): 16033.

Rosikiewicz P, Bonvin J and Sanders IR (2017) Cost-efficient production of in vitro Rhizophagus irregularis. *Mycorrhiza* 27(5): 477-486.

Savary R, Masclaux FG, Wyss T, et al. (2018) A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus Rhizophagus irregularis. *ISME J* 12(1): 17-30.

Sedlazeck FJ, Rescheneder P, Smolka M, et al. (2018) Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 15(6): 461-468.

Shannon P, Markiel A, Ozier O, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11): 2498-2504.

Smith SE and Read DJ (2008) Mycorrhizal Symbiosis, 3rd Edition. *Mycorrhizal Symbiosis, 3rd Edition*. 1-787.

Spatafora JW, Chang Y, Benny GL, et al. (2016) A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* 108(5): 1028-1046.

Spielmann M, Lupianez DG and Mundlos S (2018) Structural variation in the 3D genome. *Nat Rev Genet* 19(7): 453-467.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9): 1312-1313.

Stanke M, Diekhans M, Baertsch R, et al. (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24(5): 637-644.

Steidinger BS, Crowther TW, Liang J, et al. (2019) Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* 569(7756): 404-408.

Suzuki R and Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12): 1540-1542.

Talavera G and Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56(4): 564-577.

Tamiru M, Hardcastle TJ and Lewsey MG (2018) Regulation of genome-wide DNA methylation by mobile small RNAs. *New Phytol* 217(2): 540-546.

Tisserant E, Malbreil M, Kuo A, et al. (2013) Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc Natl Acad Sci U S A* 110(50): 20117-20122.

Trapnell C, Pachter L and Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9): 1105-1111.

van der Heijden MG, Martin FM, Selosse MA, et al. (2015) Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol* 205(4): 1406-1423.

van der Heijden MGA, Klironomos JN, Ursic M, et al. (1998) Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* 396(6706): 69-72.

Walker BJ, Abeel T, Shea T, et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11): e112963.

Wang Y, Sheng Y, Liu Y, et al. (2019) A distinct class of eukaryotic MT-A70 methyltransferases maintain symmetric DNA N6-adenine methylation at the ApT dinucleotides as an epigenetic mark associated with transcription. *Nucleic Acids Res* 47(22): 11771-11789.

Wheeler TJ and Eddy SR (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics* 29(19): 2487-2489.

Wild R, Gerasimaite R, Jung JY, et al. (2016) Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains. *Science* 352(6288): 986-990.

Wyss T, Masclaux FG, Rosikiewicz P, et al. (2016) Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus Rhizophagus irregularis. *ISME J* 10(10): 2514-2526.

Xiao CL, Zhu S, He M, et al. (2018) N(6)-Methyladenine DNA Modification in the Human Genome. *Mol Cell* 71(2): 306-318 e307.

Xie X, Lin H, Peng X, et al. (2016) Arbuscular Mycorrhizal Symbiosis Requires a Phosphate Transceptor in the Gigaspora margarita Fungal Symbiont. *Mol Plant* 9(12): 1583-1608.

Zhang B, Zhou Y, Lin N, et al. (2013) Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res* 23(9): 1522-1540.

Zhang G, Huang H, Liu D, et al. (2015) N6-methyladenine DNA modification in Drosophila. *Cell* 161(4): 893-906.

Zhang S, Lehmann A, Zheng W, et al. (2019) Arbuscular mycorrhizal fungi increase grain yields: a meta-analysis. *New Phytol* 222(1): 543-555.

Supplementary Information

Supplementary Notes

Supplementary Note 1: Genome assembly and annotation

Genomes of the six isolates were assembled into contigs and scaffolded, based on an estimated genome size of 150 Mb, with N50 scaffold lengths ranging from 543 kb in A5 to 2.4 Mb in C2 (Supplementary Table S1)[1,2]. Total scaffold counts ranged from 121 in B12 to 392 in A5, with L50 values ranging from 20 scaffolds in C2 to 61 in A5 (Supplementary Table S1). The number of assembled scaffolds did not decrease as estimated sequence coverage increased. For example, 151 scaffolds and a 20X estimated coverage in isolate A1 and 232 scaffolds with a 79X estimated coverage in isolate C3 (Supplementary Table S1). This might also be due to the difficulty in assembling genomes of dikaryotes. BUSCO analysis revealed that genome completeness ranged from 86.9% in A5 to 95.9% in C2, based on detection of complete, single-copy orthologs known to be conserved in fungi (fungi_db09; 290 orthologs) [3].

Single-copy top hit matches of annotated genes from the genome of each isolate ranged from 19071 protein hits in A5 to 20422 in A1. Phylogenetic relationships among the 6 isolates were inferred after aligning 1838 top protein hits commonly shared among all isolates, and the phylogeny confirmed previously observed results using ddRAD-seq data [4,5]. (Supplementary Figure S1a). Specifically, A5, B12 and A1 clustered separately from C2 and C5, with roughly 0.0019 substitutions per site. (Supplementary Figure S1a). Moreover, the proximity of C2 and C5 confirmed the very high degree of relatedness. Isolate C3 was the most distant with > 0.005 substitutions per site. Although this conservative approach utilized slightly less than 10% of the identified single-copy proteins in each genome, the median sequence length (log10) did not vary significantly among the isolates, ranging from 2.521 in C3 to 2.535 in C2 and C5 (Supplementary Figure S1b; Supplementary Table S2). Using a different approach, other researchers identified 8255 single-copy protein orthologs suggesting that high intraspecific variability exists among *R. irregularis* isolates [6]. Our data agree with the presence of high intraspecific variability although the estimate may, at present, be slightly inflated due to the lack of functionally validated AMF genes. This, combined with an over representation of coding sequences from *in silico* generated gene models deposited in public databases, may skew the perception of lineage-specific gene repertoires, although they exist.

Supplementary Note 2: Structural variation among *R. irregularis* isolates

The number of SVs (> 30 bp) in the genomes of the six *R. irregularis* isolates, as compared to isolate DAOM 197198, was highly variable; being much higher in isolate C3 than in A1 and A5 (Supplementary Figure S3a; Supplementary Table S3). A dendrogram, based on hierarchical clustering of the SV, was concordant with the phylogenetic distance among the isolates based on the

phylogeny based on sets of conserved proteins (Figure 1a; Supplementary Figure S3b), although differences in sequencing depth may lead to an underestimation of SV in isolates A1, A5 and B12 (Supplementary Figure S2). Insertions and deletions were much more abundant than inversions and duplications in all six isolates (Supplementary Figure S3a). Median SV length was different among isolates for each type of variation, and large SVs (> 2kb) were detected in all six isolates (Supplementary Table S4).

Repeated sequences are known to directly and indirectly trigger the formation of structural variation through a range of different mechanisms such as non-allelic homologous recombination, fork stalling and template switching or mobile element insertions. Approximately half of the SVs were found to be related to the presence of repeated sequences (Supplementary Figure S3c). In addition, the high frequency of active transposable elements (TEs) in *R. irregularis* has been suggested as a major source of genomic rearrangements and to contribute largely to its intraspecific genetic variability [6]. In order to address this, the occurrence of the breakpoints of SVs in known families of TEs was investigated. Similar results were observed across the six isolates, where DNA transposons and simple repeats were more commonly associated with the presence of SVs (Supplementary Figure S3d, Supplementary Table S5). These results provide a first insight into the contribution of TEs to the existence of large genomic variations among *R. irregularis* isolates.

Unsurprisingly, most of the SV was found to be located in non-coding regions (from 92% in A1, to 96% in C2). However, in this study, the number of genes directly affected by the presence of SVs compared to the model *R. irregularis* isolate DAOM 197198 ranged from 1517 genes in isolate A1 to 13081 genes in isolate C3 (Supplementary Figure S3e; Supplementary Table S6). Gene ontology (GO) enrichment analysis showed that some biological processes and molecular functions were overrepresented in genes harbouring SV in isolates B12, C2, C5, and C3 (Supplementary Data 1). More specifically, processes such as signalling, protein phosphorylation, regulation of P metabolic processes and activation of the MAPK cascade were significantly enriched in isolates C2, C5 and C3 (Supplementary Figure 3f). Given that these are all involved in P transport and pathways associated with P regulation, this suggests that structural variation can affect regulation of symbiotic effects.

Supplementary Note 3: Structural variation between two *R. irregularis* clones (isolates C2 and C5)

No clear differences could be observed in SV between isolates C2 and C5. Supplementary Figure S3 shows the negligible number of SVs detected between C2 and C5, when using DAOM 197198 as the reference assembly. Mapping reads of C2 and C5 to the *de novo* C2 genome assembly further reduced the number of SVs between C2 and C5. While it is inevitable that some SV will be detected because of sequencing issues or assembly issues, after manually curating putative isolate-specific variants against alignment files in IGV, it was clear that almost all these variants were differentially called as a result of low read support values and/or insufficient coverage. These were, therefore, disregarded (an example is shown in Supplementary Figure S4). Coupled with the fact that Wyss et al. (2016) could not find convincing SNP differences between these two isolates, these results support the hypothesis that these two *R. irregularis* isolates are almost certainly clones as they are genetically indistinguishable.

Supplementary Note 4: Structural variation between nuclei

Structural variation existing within the dikaryon isolate C3 most likely represents structural differences between its two nucleus genotypes. The amount of structural variation (7732 SVs) fell within the range of variation observed between some of the genetically different isolates and isolate DAOM197198 Supplementary Figure 5a). Assuming that this represents SV differences between the two nucleus genotypes, this suggests that some between nucleus differences are as large as SV between some *R. irregularis* isolates. An alternative hypothesis would be that recombination among nuclei, as previously observed in one *R. irregularis* isolate, could lead to within-fungus SV among nuclei [7]. However, no convincing evidence for recombination was observed between nuclei in isolate A4 [7], a clone of C3, making this alternative hypothesis unlikely. Structural variation within C3 exhibited similar properties to those observed among the isolates. The vast majority of SVs were either insertions (3132) or deletions (4436) (Supplementary Figure S5a) and a large proportion were related to the presence of repeated sequences (3426 SVs) (Supplementary Figure S5b). Additionally, DNA transposons were more associated with SVs than other known TE families (Supplementary Figure S5c). Finally, SV between the two nucleus genotypes occurred in a remarkable number of genes (2437) (Supplementary Figure S5d), although no significant GO terms were enriched. Together, these results suggest that genomic variation occurring among nuclei within the dikaryon isolate C3 is as large, and shares similar characteristics, to that observed between some homokaryote *R. irregularis* isolates.

Supplementary Note 5: 6mA epigenetic signatures differ between nuclei in dikaryons

We analysed whether the two nucleus genotypes in the dikaryon isolate C3 were differentially methylated. We were only able to perform this analysis on gene sequences that we knew were located on different nuclei. Thus, this restricted our analysis to genes that revealed structural differences between the two nucleus genotypes. Of these, 331 genes were methylated. 153 of them lacked 6mA ApT marks in either primary assembly or in the secondary haplotig. The absence of methylation on genes from one nucleus genotype, and the presence in the other, indicates that differential epigenetic gene regulation can occur between nucleus genotypes in dikaron *R. irregularis*. Because we were restricted to only identifying differential methylation occurring in genes between the two nucleus genotypes, at selected sites that differed in their DNA sequence, this represents a very small sub-sample of the total number of methylated genes in the genome. However, at those sites, 46% showed differential methylation suggesting that levels of methylation differences between the two nucleus genotypes are very high.

Supplementary Note 6: The majority of AMF genes involved in P metabolism, P sensing and signalling and homeostasis are under 6mA epigenetic regulation

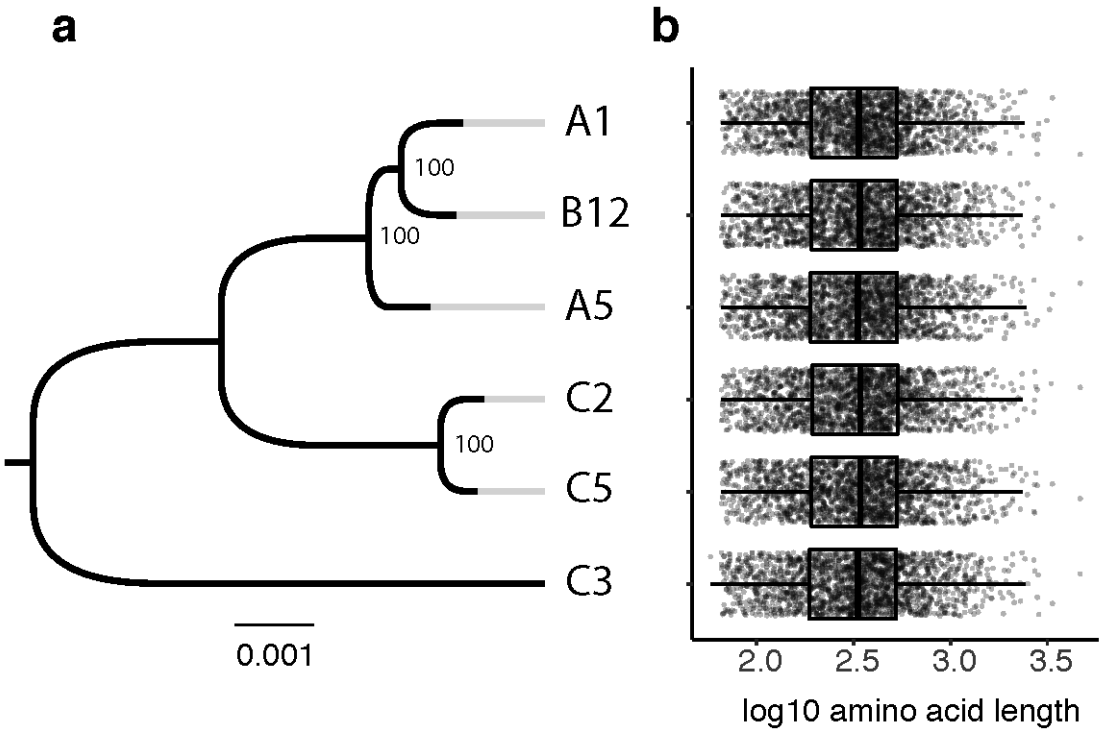
We performed a targeted assessment of SV and 6mA presence/absence in genes involved in P transport, P metabolism, signalling and homeostasis, P responsive signalling, as well as genes in other pathways shown to participate in P homeostasis in S*accharomyces cerevisiae* or in AMF [8-12]. We also included genes from the PKA, MAPK and Tor signalling pathways that are influenced by changes in inorganic P supply to AMF. Furthermore, we also included both high and low affinity sugar transporters as they have been shown to be associated expression of high-affinity P transporters [13].

We found that high-affinity P transporters (PT3, PT4) and sugar transporters (MST2, MST3) harboured SV but were not methylated (Supplementary Table 7). On the contrary, the low-affinity SPX domain containing P transporters (PHO87, PHO88, PHO90, PHO91) and the low-affinity sugar transporter SUC1 were methylated in all three isolates and MST4 was methylated in two isolates. SPX domain proteins have been suggested to be key in P transport and metabolism in eukaryotes including AM fungi [9,14]. Remarkably, all the genes that have been identified in previous studies to be involved in inositol polyphosphate synthesis/hydrolysis and P responsive signalling were 6mA methylated, thus, appearing to be under epigenetic control. Such genes are thought to be very important in AMF because they have to take up and regulated levels of P that are much greater than that needed by the fungus alone. The vacuolar transporter chaperones VTC1, VTC 2 and VTC4, involved in polyphosphate synthesis, and which also contain the SPX domain, were also methylated with some differences between in isolate C5 and its clone C2, suggesting isolate specific epigenetic differences of this methylation mark in fine tuning polyphosphate accumulation. Finally, the parts of the of the PKA signalling pathways, that are upregulated in AMF in response to low Pi environments, and the MAPK and Tor signalling pathways, that are down-regulated at low Pi appear to be almost completely

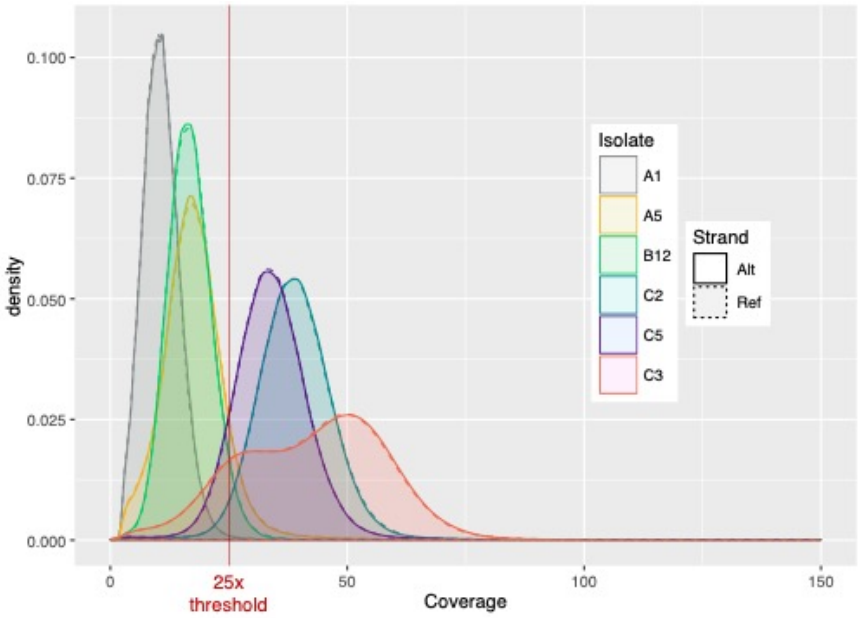
methylated, suggesting strong epigenetic control on the regulation of these pathways in the fungus in varying P environments.

Given the critical role of AMF in P transport to plants, the strong methylation of genes involved in almost all components of polyphosphate biosynthesis and release, and P-responsive signalling suggests 6mA plays an important role in P homeostasis. The majority of the 6mA harbouring genes were methylated in all three isolates. Despite this, a few genes exhibited differential methylation among the isolates which may yield a big impact on either on AMF physiology or differential plant growth. Compared to methylation, the occurrence of SV was far less frequent, especially in gene body which is consistent with an expected high conservation of the P transport and metabolism related genes in this species.

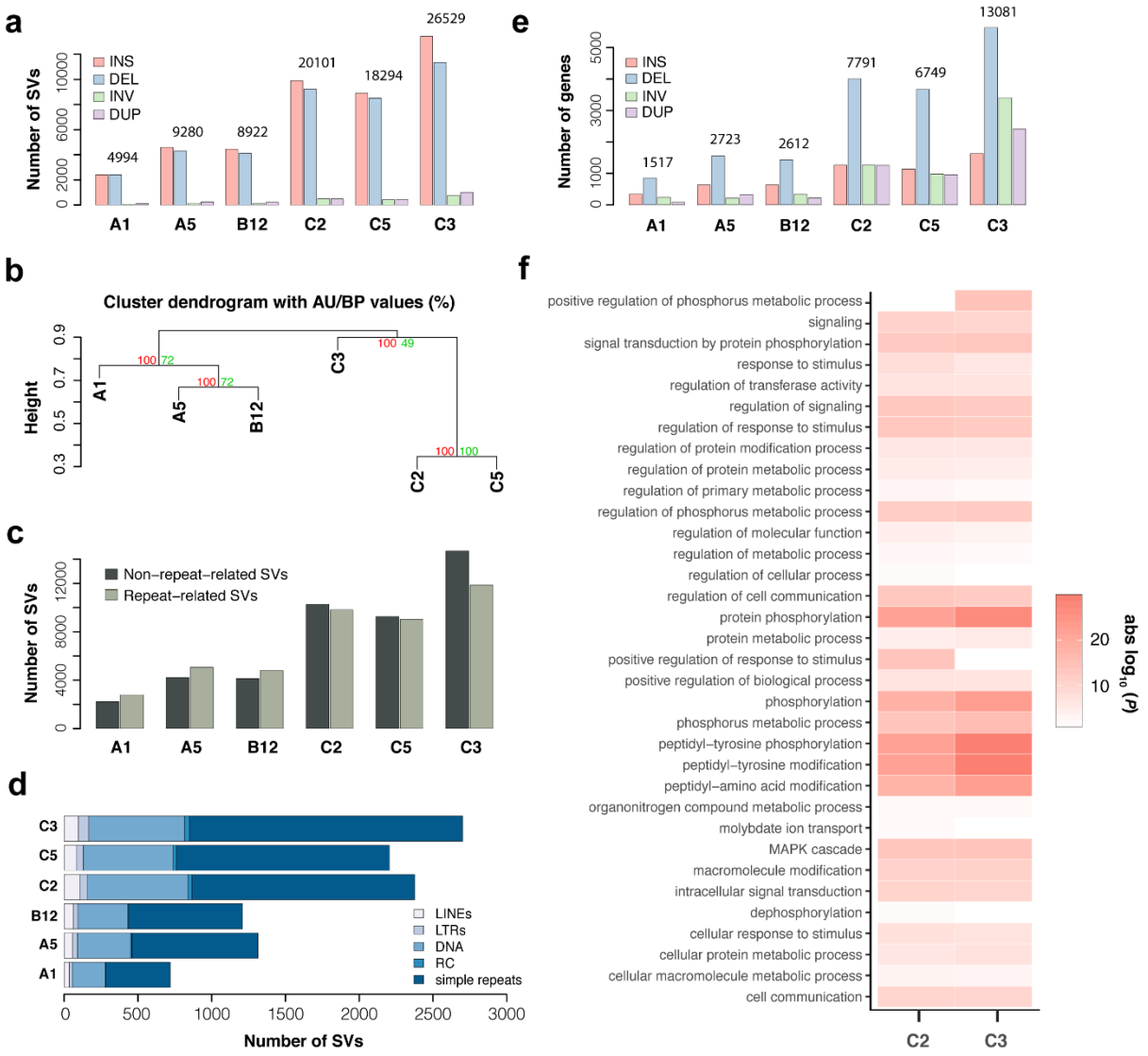
Supplementary Figures



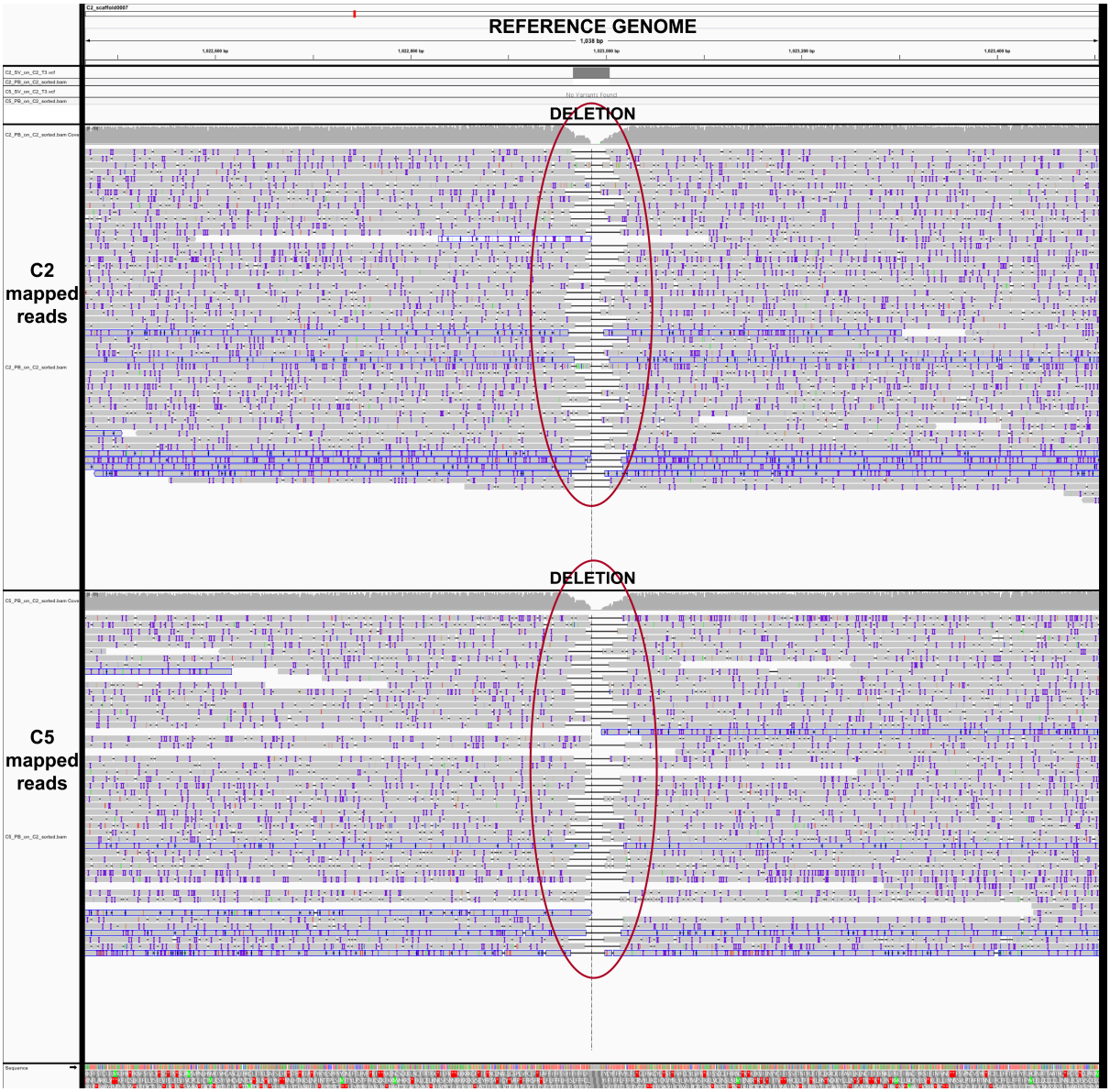
Supplementary Figure S1. Phylogeny and sequence lengths of conserved genes of six *Rhizophagus irregularis* isolates. (a) Phylogenetic distance of concatenated single copy UniProt accessions (n=1838) of *R. irregularis* isolates and computed by bootstrapping 1000 iterations with RAxML. The isolates A5, B12 and A1 clustered separately from C2 and C5, with roughly 0.0019 substitutions per site while proximity of C2 and C5 confirmed the very high degree of relatedness, whereas Isolate C3 was the most distant with > 0.005 substitutions per site **(b)** Box plots of individual amino acid lengths concatenated for alignment showing that the median sequence length (log10) did not vary significantly among the isolates, ranging from 2.521 in C3 to 2.535 in C2 and C5 (Supplementary Table 2).



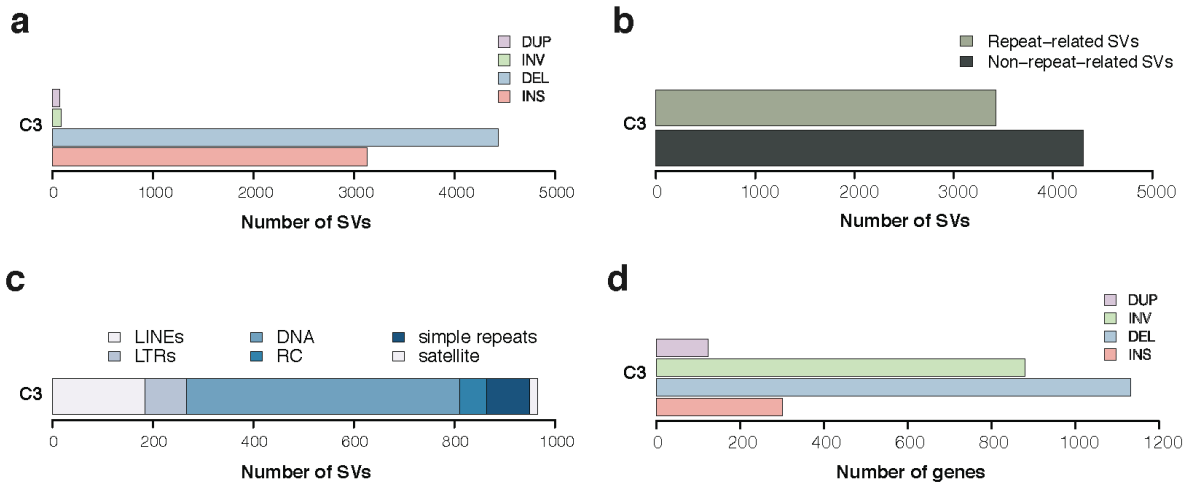
Supplementary Figure S2. Sequence read coverage distribution per DNA strand of six *R. irregularis* isolates. The 25x coverage/strand threshold is considered necessary for reliable 6mA calling.



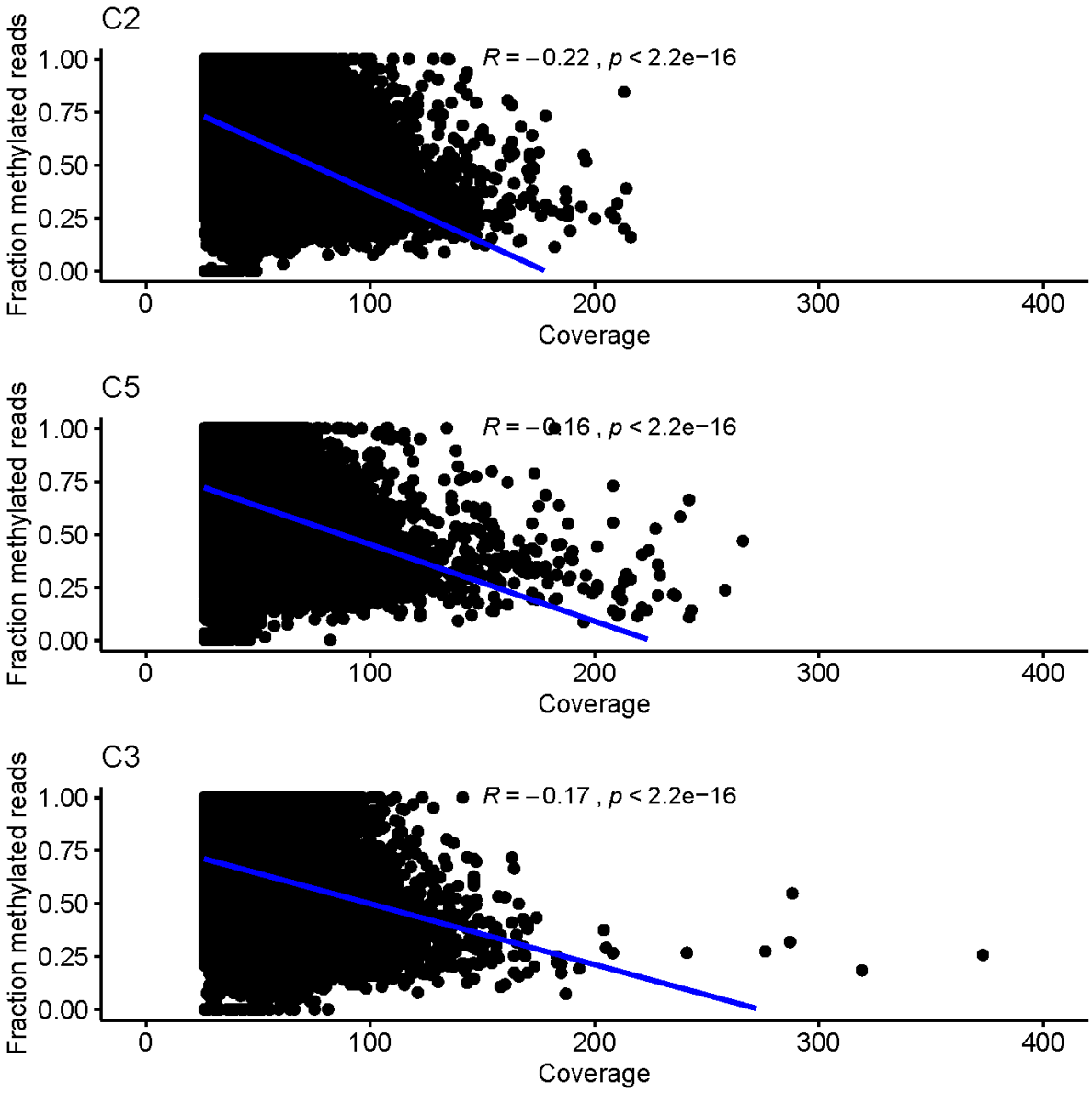
Supplementary Figure S3. Structural variation among genomes of six *R. irregularis* isolates compared to the *R. irregularis* DAOM197198 genome. **(a)** Number of SVs detected in the six genomes by SV type (INS: insertion; DEL: deletion; INV: inversion; DUP: duplication). **(b)** Hierarchical clustering of the six isolates based on the number of co-occurring SVs (AU: approximately unbiased probability shown in red; BP: bootstrap probability shown in green). **(c)** Number of repeat-related and non-repeat related SVs. **(d)** Number of SVs associated with known families of TEs (LINEs: long interspersed nuclear elements; LTRs: long terminal repeats retrotransposons; DNA: DNA transposons; RC: rolling circle transposons). **(e)** Number of genes where SV occurred. **(f)** Enriched GO categories assessed in the genes affected by SV in isolates C2 and C3. Define abs log 10 P.



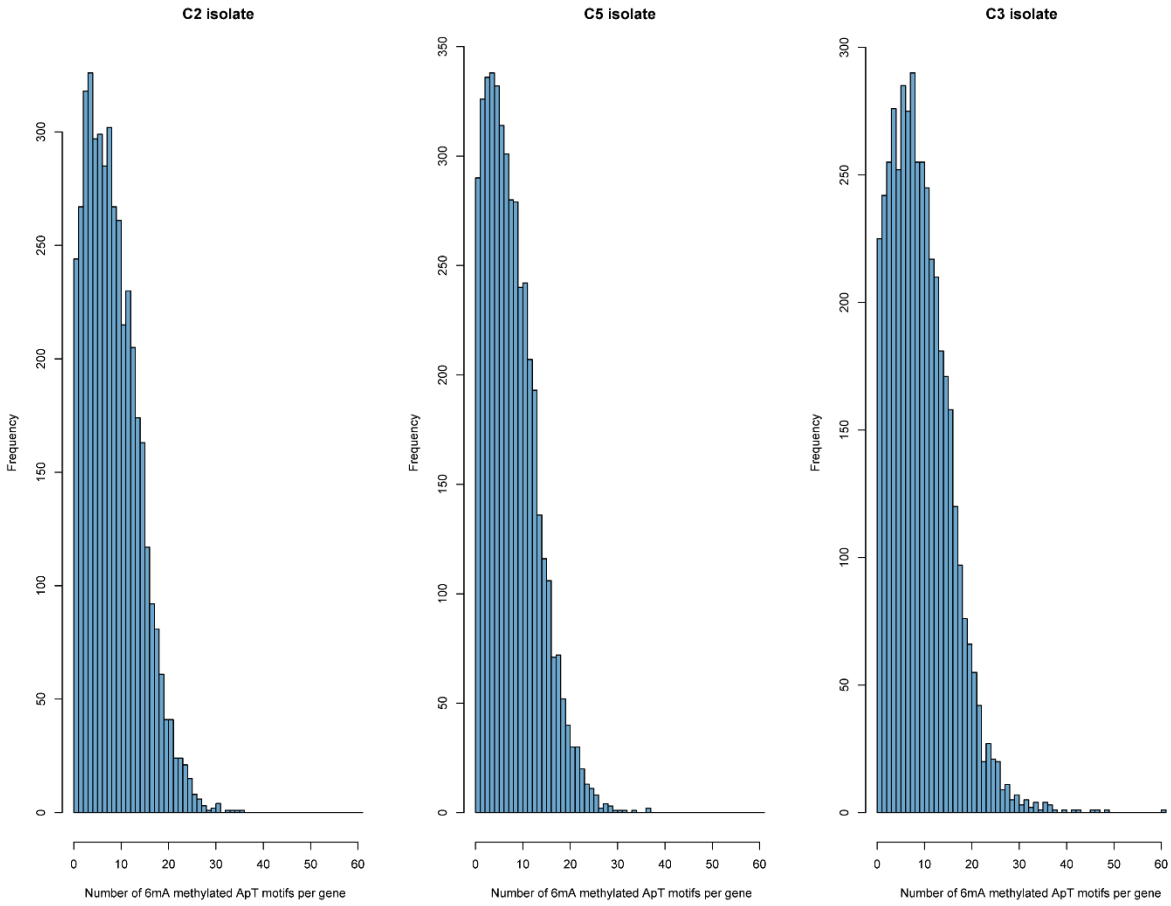
Supplementary figure S4. An example of the visualization of a discordant deletion between isolates C2 and C5 due to the lack of read support in C5. The deletion is clearly observed in the mapped reads of both isolates (a sudden drop of coverage in the highlighted genomic region). This SV was initially called in both isolates, but the read support value (RE) in isolate C5 was lower than a quality-filtering threshold (RE >15), while it was higher in isolate C2 (RE=18). Such sites were discarded after visual checks not considered as a structural variant between isolates C2 and C5.



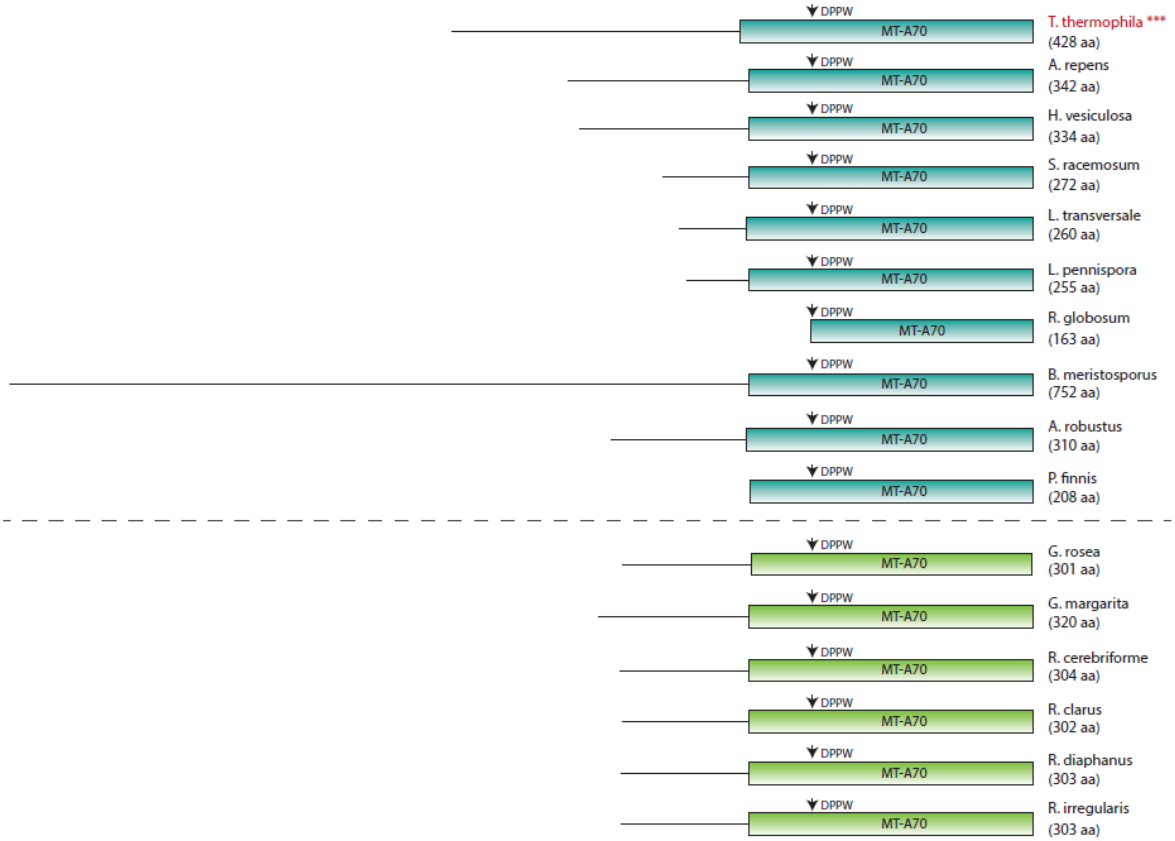
Supplementary Figure S5. **Structural variation (SV) within the dikaryon *R. irregularis* isolate C3. (a)** Number of SVs detected between the two nucleus genotypes. **(b)** Number of SVs related and non-related with the presence of repeated sequences. **(c)** Number of SVs associated with known families of TEs (LINEs: long interspersed nuclear elements; LTRs: long terminal repeats retrotransposons; DNA: DNA transposons; RC: rolling circle transposons). **(d)** Number of genes from the primary assembly that contained SVs in the haplotig.



Supplementary figure S6. The relationship between the fraction of methylated to non-methylated reads and coverage in three *R. irregularis* isolates. The Pearson correlation coefficients showed a very weak negative linear relationship despite being significant in all three isolates, suggesting that coverage had a negligible effect on heterogeneity of 6mA methylation in the dataset.



Supplementary figure S7. Distribution of number of 6mA methylated ApT motif per gene in 4398 commonly methylated genes in three *R. irregularis* isolates C2, C5 and C3.



Supplementary figure S8. Conserved MT-A70 domain and catalytic motif (DPPW) in putative AMT1 MTases of EDF species. The arrow indicates the position of the DPPW motif within the sequence. The length of the protein sequences are indicated in brackets (number of aminoacids). *T. thermophila* AMT1 is shown as a reference (colored in red and marked with three asterisks). The dashed line separates the Glomeromycotina species (green) from other EDF species (blue).

Supplementary tables

Supplementary Table S1. Genome assembly and annotation statistics of six *R. irregularis* isolates Define what BUSCO completeness, fragmented and missing means.

	A1	**A5**	**B12**	**C2**	**C3**	**C5**
PacBio SMRT cells (RS II)	6	6	7	12	24	11
Number of scaffolds	151	392 (428 haplotigs)	121	123	232 (1084 haplotigs)	372
Assembled genome size (Mb)	147.98	115.47	146.58	159.8	159.63	166
Scaffold N50	25	61	25	20	30	36
GC%	28.09	27.39	28.09	28.5	27.94	27.64
N%	0.233	1.117	0.139	0.255	3.212	1.387
Coverage	20.52X	36X	30X	64X	79X	62X
Gene number	26293	28267	25711	27251	25670	27461
BUSCO completeness % **	92.1	86.9	95.9	95.9	95.1	95.5
BUSCO fragmented % **	4.1	3.8	2.4	2.1	3.1	2.1
BUSCO missing % **	3.8	9.3	1.7	2.0	1.8	2.4

Supplementary Table S2. One-way analysis of variance and Tukey multiple comparisons of means between isolates for log10(lengths) of proteins. 95% family-wise confidence level.

	Df	SumSq	MeanSq	Fvalue	Pr(>F)
log10(lengths) of proteins	5	0.1	0.0249	0.234	0.948
residuals	11022	1171.0	0.1062		

Comparison isolates (log10(lengths) of proteins)	**Difference**	**Lower**	**Upper**	**Probability**
A5-A1	-0.00302	-0.03367	0.027626	0.999766
B12-A1	0.000303	-0.03034	0.030949	1
C2-A1	0.002051	-0.02859	0.032697	0.999965
C3-A1	-0.00826	-0.03891	0.022384	0.972788
C5-A1	-0.00024	-0.03088	0.030409	1
B12-A5	0.003323	-0.02732	0.033969	0.999626
C2-A5	0.005071	-0.02557	0.035718	0.997108
C3-A5	-0.00524	-0.03589	0.025404	0.996616
C5-A5	0.002783	-0.02786	0.033429	0.999843
C2-B12	0.001749	-0.0289	0.032395	0.999984
C3-B12	-0.00856	-0.03921	0.022081	0.968181
C5-B12	-0.00054	-0.03119	0.030107	1
C3-C2	-0.01031	-0.04096	0.020332	0.930671
C5-C2	-0.00229	-0.03293	0.028358	0.99994
C5-C3	0.008025	-0.02262	0.038671	0.97605

Supplementary table S3. Number of structural variations (SV) detected in six *R. irregularis* isolates as compared to isolate DAOM 197198.

	Number of structural variation events				
Isolate	**INSERTION**	**DELETION**	**INVERSION**	**DUPLICATION**	**TOTAL**
A1	2413	2413	57	111	4994
A5	4592	4308	126	254	9280
B12	4455	4119	123	225	8922
C2	9902	9226	485	488	20101
C5	8907	8518	425	444	18294
C3	13436	11347	750	996	26529

Supplementary table S4. The minimum, median and maximum lengths of each type of structural variant (SV) in the six *R. irregularis* isolates.

	Structural variant length											
	INSERTION			**DELETION**			**INVERSION**			**DUPLICATION**		
Isolate	**Min.**	**median**	**Max.**	**Min.**	**median**	**Max.**	**Min.**	**median**	**Max.**	**Min.**	**median**	**Max.**
A1	31	251	7930	31	352	177719	85	2078	837238	59	558	105948
A5	31	239	11244	31	410,5	304960	72	1108	237907	34	594,5	308333
B12	31	247	12094	31	363	104869	75	712	837238	51	555	231310
C2	31	239	22770	31	394	867212	66	1292	237908	35	868	871321
C5	31	249	13553	31	420	867212	71	1143	197265	35	841	871321
C3	31	196	14501	31	371	572715	69	2495,5	575919	37	2012	575869

Supplementary table S5. Number of structural variants (SV) related to the presence of transposable elements (TE) in the six *R. irregularis* isolates.

Number of transposable elements -enabled structural variants								
Isolate	**LINEs***	**LTRs***	**DNA***	**RC**	**Simple_repeat**	**Low_complexity**	**Unknown**	**Total**
A1	35	20	223	3	438	67	1445	2231
A5	58	32	361	8	855	124	2478	3916
B12	61	32	337	7	772	109	2437	3755
C2	108	51	681	26	1511	231	4812	7420
C5	83	47	609	22	1444	221	4397	6823
C3	96	70	649	34	1852	278	5361	8340

*LINEs: Long interspersed nucleotide elements; LTRs: Long terminal repeats; RC: Rolling Circle

Supplementary table S6. Number of genes directly containing structural variation (SV) in the six *R. irregularis* isolates.

	Number of genes				
Isolate	**INSERTION**	**DELETION**	**INVERSION**	**DUPLICATION**	**TOTAL**
A1	342	856	237	82	1517
A5	633	1553	212	325	2723
B12	635	1426	336	215	2612
C2	1261	4007	1266	1257	7791
C5	1135	3673	983	958	6749
C3	1631	5638	3402	2410	13081

Supplementary table S7. Characterization of structural variation and 6mA ApT epigenetic marks occurring in genes involved in P transport, P metabolism, P responsive signalling, inositol polyphosphate biosynthesis and hydrolysis, PKA signalling, MAPK signalling, Tor signalling and sugar transport genes in *R. irregularis*. in isolates C2, C5 and C3. Isolates C2 and C5 are genetically indistinguishable. E: Exon; I: Intron and D: Downstream. Green represents the presence of structural variation and 6mA epigenetic marks and pink represents their absence.

Genes	Structural variation		6mA		
	C2	C3	C2	C5	C3
Phosphate Transport					
PT1					
PT2					
PT3	D	D			
PT4	E	E			
PT5					
PT6					
PHO87					
PHO90					
PHO91					
HA5					
Pho88					
Phosphate Metabolism					
Pho5					
ALP					
Vtc1					
Vtc2					
Vtc4	U	U			
Ppn1		D			
Gde1	D	D			
Vma10					
Phosphate responsive signalling/Inositol polyphosphate biosynthesis/hydrolysis					
VIP1	U	U			
PLC1					
ARG82					
KCS1	U				
IPK1					
DDP1		D			
Pho81					
Pho80	D	D			
Pho85					
PKA signalling					
TPK1					
Sch9	D	D			
NTH1		D			
TPS2					
TSL1	D	D			
Rim15					
Msn4					
Gis1	D				
SSA3		U			
HSP20					
SOD1	U	U			
SOD2					
BCY1					
MAPK signalling					
Mapk2					
Mek-2					
Mek-1					
NRC-1					
os-4		D			
os-5					
os-2					
Mik-1					
Mkc-1					
Tor signalling					
Tor2		I			
Gad8					
ste20					
Sin1					
Mip1					
Sugar Transport					
MST2	E, D	E, D			
MST3					
MST4					
SUC1					
E: Exon; I: Intron; U: Upstream; D: Downstream	Presence	Absence			

Supplementary notes References

1 Sedzielewska, K. A. *et al.* Estimation of the Glomus intraradices nuclear DNA content. *New Phytol* **192**, 794-797 (2011).

2 Tisserant, E. *et al.* Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc Natl Acad Sci U S A* **110**, 20117-20122 (2013).

3 Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* (2017).

4 Savary, R. *et al.* A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus Rhizophagus irregularis. *ISME J* **12**, 17-30 (2018).

5 Wyss, T. *et al.* Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus Rhizophagus irregularis. *ISME J* **10**, 2514-2526 (2016).

6 Chen, E. C. H. *et al.* High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. *New Phytol* **220**, 1161-1171 (2018).

7 Chen, E. C. *et al.* Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. *Elife* **7** (2018).

8 Azevedo, C. & Saiardi, A. Eukaryotic Phosphate Homeostasis: The Inositol Pyrophosphate Perspective. *Trends Biochem Sci* **42**, 219-231 (2017).

9 Ezawa, T. & Saito, K. How do arbuscular mycorrhizal fungi handle phosphate? New insight into fine-tuning of phosphate metabolism. *New Phytol* **220**, 1116-1121 (2018).

10 Garcia, K. *et al.* Take a Trip Through the Plant and Fungal Transportome of Mycorrhiza. *Trends Plant Sci* **21**, 937-950 (2016).

11 Xie, X. *et al.* Arbuscular Mycorrhizal Symbiosis Requires a Phosphate Transceptor in the Gigaspora margarita Fungal Symbiont. *Mol Plant* **9**, 1583-1608 (2016).

12 Eskes, E., Deprez, M. A., Wilms, T. & Winderickx, J. pH homeostasis in yeast; the phosphate perspective. *Curr Genet* **64**, 155-161 (2018).

13 Helber, N. *et al.* A versatile monosaccharide transporter that operates in the arbuscular mycorrhizal fungus Glomus sp is crucial for the symbiotic relationship with plants. *Plant Cell* **23**, 3812-3823 (2011).

14 Wild, R. *et al.* Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains. *Science* **352**, 986-990 (2016).

Chapter 3

Generation of disproportionate nuclear genotype proportions in *Rhizophagus irregularis* progeny causes allelic imbalance in gene transcription

Chanz Robbins*[1], Joaquim Cruz Corella*[1], Consolée Aletti[1], Réjane Seiler[1], Ivan D. Mateus[1], Soon-Jae Lee[1], Frédéric G. Masclaux[2], Ian R. Sanders#[1]

[1] Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland

[2] Group of Genetic Medicine, Geneva University Hospital, Building D, 1205 Geneva, Switzerland

* These authors contributed equally to this work

Corresponding author: Ian R. Sanders

This chapter was published in *New Phytologist* in June 4th, 2021.

Abstract

Arbuscular mycorrhizal fungi (AMF) form mutualisms with most plant species. The model AMF *Rhizophagus irregularis* is common in many ecosystems and naturally forms homokaryons and dikaryons. Quantitative variation in allele frequencies in clonally dikaryon offspring suggests they disproportionately inherit two distinct nuclear genotypes from their parent. This is interesting because such progeny strongly and differentially affect plant growth. Neither the frequency and magnitude of this occurrence, nor its effect on gene transcription, are known. Using reduced representation genome sequencing, transcriptomics and quantitative analysis tools, we show that progeny of homokaryons and dikaryons are qualitatively genetically identical to the parent. However, dikaryon progeny differ quantitatively due to unequal inheritance of nuclear genotypes. Allele frequencies of actively transcribed bi-allelic genes resembled the frequencies of the two nuclear genotypes. More bi-allelic genes showed transcription of both alleles than mono-allelic transcription but bi-allelic transcription was less likely with greater allelic divergence. Mono-allelic transcription levels of bi-allelic genes were reduced compared to bi-allelic gene transcription; a finding consistent with genomic conflict. Given that genetic variation in *R. irregularis* is associated with plant growth, our results establish quantitative genetic variation as a future consideration when selecting AMF lines to improve plant production.

Keywords: AMF genetics; Arbuscular mycorrhiza; plant symbiosis; plant production; *Rhizophagus irregularis*

Introduction

Arbuscular mycorrhizal fungi (AMF; Glomeromycotina) are ubiquitous soil microorganisms that establish mutualistic relationships with most terrestrial plants (Davison et al., 2015; Van der Heijden et al., 1998; Brundrett and Tedersoo, 2018). Hyphae of these fungi absorb and transport soil inorganic nutrients, especially phosphorus and nitrogen, to plant roots (Fellbaum et al., 2012; Govindarajulu et al., 2005). In exchange, AMF receive photoassimilates and plant-derived lipids (Bravo et al., 2017; Keymer et al., 2017). This symbiotic interaction occurs across the planet, making AMF global players in nutrient and carbon cycling, affecting plant growth and diversity (Bago et al., 2000; Van der Heijden et al., 1998; Steidinger et al., 2019). Accordingly, *Rhizophagus irregularis* is a model AMF species of Glomeracae; a dominant family of AMF communities (Gao et al., 2019; Rodriguez-Echeverria et al., 2017; Tisserant et al., 2013).

AMF are coenocytic, with thousands of nuclei coexisting within a common cytoplasm. Recent studies show that *R. irregularis* isolates are either homokaryons or dikaryons, and that nuclei are haploid (Chen et al., 2018b; Masclaux et al., 2018; Masclaux et al., 2019; Ropars et al., 2016; Wyss et al., 2016). Dikaryon AMF harbour populations of two genetically distinct nuclei; referred to hereon as nuclear genotypes (Masclaux et al., 2018). Although evidence implies that *R. irregularis* might reproduce sexually, population genetic studies suggest that clonal reproduction occurs frequently in nature (Mateus et al., 2020; Ropars et al., 2016; Savary et al., 2018a). Indeed, *R. irregularis* isolates have been maintained clonally for almost 20 years *in-vitro* (Koch et al., 2004; Rosikiewicz et al., 2017). Since initiating cultures, dikaryon isolates continually retain both nuclei (Angelard et al., 2010; Masclaux et al., 2018; Masclaux et al., 2019; Ropars et al., 2016; Wyss et al., 2016).

Single-spore culturing is a technique that allows the generation of AMF single spore sibling lines (SSSLs) from an AMF isolate that we call here a parental isolate (Fig. 1a). At present, unequal inheritance of nuclear genotypes has only been described for a small number of SSSLs from one parental isolate, known as C3 (Angelard et al., 2014; Croll et al., 2009; Ehinger et al., 2012; Masclaux et al., 2018). Three studies demonstrated that SSSLs of C3 (with an approximate 1:1 ratio of the two nuclei) can inherit different proportions of both nuclear genotypes (Angelard et al., 2010; Angelard et al., 2014; Masclaux et al., 2018). The first study detected genetic variation among SSSLs by assessing amplified fragment length polymorphisms (AFLPs) (Angelard et al., 2010). As discussed by Angelard et al. (2010), although this multi-locus technique can verify the presence of an allele, it lacks sufficient sensitivity to measure changes in allele frequency or among SSSLs and doesn't distinguish single copy from multicopy regions of the genome. Both capillary electrophoresis, and amplicon sequencing at a single locus (known as the *bg112* locus), confirmed that nuclear genotypes can be unequally inherited among SSSLs of C3 (Angelard et al., 2010; Ehinger et al., 2012; Masclaux et al., 2018). These results indicated that, significant shifts in allele frequencies sometimes arise among clonally produced dikaryon spores. The study by Masclaux et al. (2018) considered one single-copy locus in few SSSLs, where PCR error represents an unlikely, but possible, source of variation. A high-throughput, reduced-

representation approach, such as double-digest restriction site associated DNA sequencing (ddRAD-seq), is a reliable method for detecting genetic diversity within populations and distinguishing quantitative differences in allele frequencies (Wyss et al., 2016).

Whole-genome amplification and sequencing of individual nuclei were fundamental in establishing that *R. irregularis* isolates have either a homokaryon or dikaryon genome organization (Chen et al., 2018a; Lin et al., 2014; Ropars et al., 2016). Yet, this technique offers too low resolution and, thus, is not well suited for determining quantitative differences in allele frequencies among several lines. First, the success rate of obtaining data of sufficient quality from a single-nucleus of *R. irregularis* is staggeringly low, ranging from 10-63% (Chen et al., 2018a; Lin et al., 2014; Ropars et al., 2016). Second, the quantity of data needed to address quantitative variation at multiple loci among several lines, is fiscally prohibitive. For example, to date, less than 300 nuclei have been sequenced, of which, only 148 passed quality filters (4, Lin et al. (2014); 59, Ropars et al. (2016); 85 Chen et al. (2018a), respectively). A conservative assessment of changes in nuclear dynamics among dikaryon SSSLs would require a minimum of 1000 successfully sequenced nuclei from each SSSL. For this reason, ddRAD-seq is more suitable to estimate quantitative genetic variation existing among SSSLs at multiple bi-allelic sites across the genome. By using ddRAD-seq, allele frequency variation can be estimated in many dikaryon and homokaryon SSSLs and compared to their parent to quantify changes in nuclear dynamics. The premise of this analysis rests on the fact that bi-allelic sites must be single-copy regions of the genome, meaning that detection of two alleles would only be possible if two different nuclear genotypes were represented. Thus, estimating the frequency of the two alleles can serve as a proxy for the relative abundance of both nuclear genotypes in a dikaryon. It is true that some bi-allelic sites were still detected in ddRAD-seq and whole genome sequencing of homokaryons (Masclaux et al., 2019; Savary et al., 2018b; Wyss et al., 2016; Chen et al., 2018b). However, these loci are very few, located in problematic regions of the assembly, and seem to have no discernible functional consequence (Masclaux et al., 2019). In stark contrast, bi-allelic sites in dikaryon C3 were more prevalent, and likely impact bi-allelic gene expression (Masclaux et al., 2019).

Quantitative genetic variation among SSSLs is likely significant for their symbiotic interaction with plants since SSSLs differ significantly in fungal quantitative traits, how they colonize roots and how they affect plant biomass (Angelard et al., 2014; Ehinger et al., 2012; Savary et al., 2018a). Indeed, pot experiments with rice, as well as field studies with cassava, indicate that genetic variation among SSSLs has enormous effects on plant biomass (Angelard et al., 2010; Ceballos et al., 2019; Ceballos et al., 2013; Mateus et al., 2019; Savary et al., 2020). The link between qualitative genetic variation (presence or absence of single nucleotide polymorphisms; SNPs) of *R. irregularis* isolates and plant growth was recently presented by Ceballos et al. (2019), although likely depends additionally on plant host, edaphic characteristics and other biotic and abiotic factors. However, the more elusive link between quantitative genetic variation (allele proportions) among SSSLs and its effect on plant growth has not yet been made. It is first necessary to understand whether quantitative changes in nuclear dynamics lead to quantitative differences in gene expression of dikaryon SSSLs.

How often, and by how much, quantitative differences in allele frequencies vary among dikaryon SSSLs could have profound consequences on fungal gene expression and on the AMF-plant symbiosis. For example, imbalanced nuclear ratios affect gene transcription and growth rate of the heterokaryon basidiomycete, *Heterobasidion parviporum*, resulting in phenotypic differences from true diploid individuals (Clergeot et al., 2019). Moreover, due to intricacies of transcriptional regulation within nuclei, equal proportions of two nuclei may not necessarily result in equal allele transcription. This may be due to localized transcriptional bursts, allele-specific gene imprinting, or other mechanisms (Dong et al., 2017; Lafon-Placette et al., 2018; Larsson et al., 2019). For example, the dikaryon basidiomycete *Agaricus bisporus*, exhibits imbalanced allele expression at different growth stages, despite both nuclear genotypes being equally abundant (Gehrmann et al., 2018). Interestingly, many bi-allelic sites in the dikaryon *R. irregularis* isolate, C3, were expressed in proportions equal to nuclear genotype proportions estimated from ddRAD-seq data, as well as the frequencies of both *bg112* alleles (Masclaux et al., 2018). Although this study showed that both nuclei were transcriptionally active in dikaryons, it could not address the effects of unbalanced nuclear dynamics on the contribution of gene expression in SSSLs from each of the two different nucleus genotypes. To address that, gene transcription needs to be assessed among SSSLs that have variable proportions to the two nuclear genotypes. Our current knowledge of how frequent nuclear genotype proportions vary in *R. irregularis*, as well as consequences on transcription, is very limited (Kokkoris et al., 2020; Yildirir et al., 2020; Kokkoris et al., 2021).

To assess quantitative variation at bi-allelic sites, 48 SSSLs were generated from 3 homokaryon and 2 dikaryon 'parental' isolates of *R. irregularis* (Fig. 1a,b). The parental isolates represent single spore cultures from a field in Switzerland, and have been propagated clonally in axenic conditions for almost 20 years (Koch et al., 2004). We employed ddRAD-seq to study allele frequencies at hundreds of bi-allelic sites to test the prevalence and amplitude of quantitative genetic variation among SSSLs (Fig. 1b,c). We later focused on six dikaryon SSSLs of C3 to investigate allelic-imbalance in expressed genes at single-copy bi-allelic sites, and whether allele proportions reflect nuclear dynamics (Fig. 1c). Here, we define allelic-imbalance to mean unequal transcription of two alleles of single-copy bi-allelic genes, such that the two alleles are located on different nuclear genotypes and transcription of each allele represents the transcriptional contribution of each nuclear genotype. We hypothesized that transcribed alleles at bi-allelic sites would reflect DNA allele frequencies detected with ddRAD-seq. We investigated allelic expression patterns across hundreds of bi-allelic sites to further test whether all bi-allelic genes in an *R. irregularis* dikaryon displayed bi-allelic expression (*i.e.,* both copies expressed), or if some genes only exhibited mono-allelic expression.

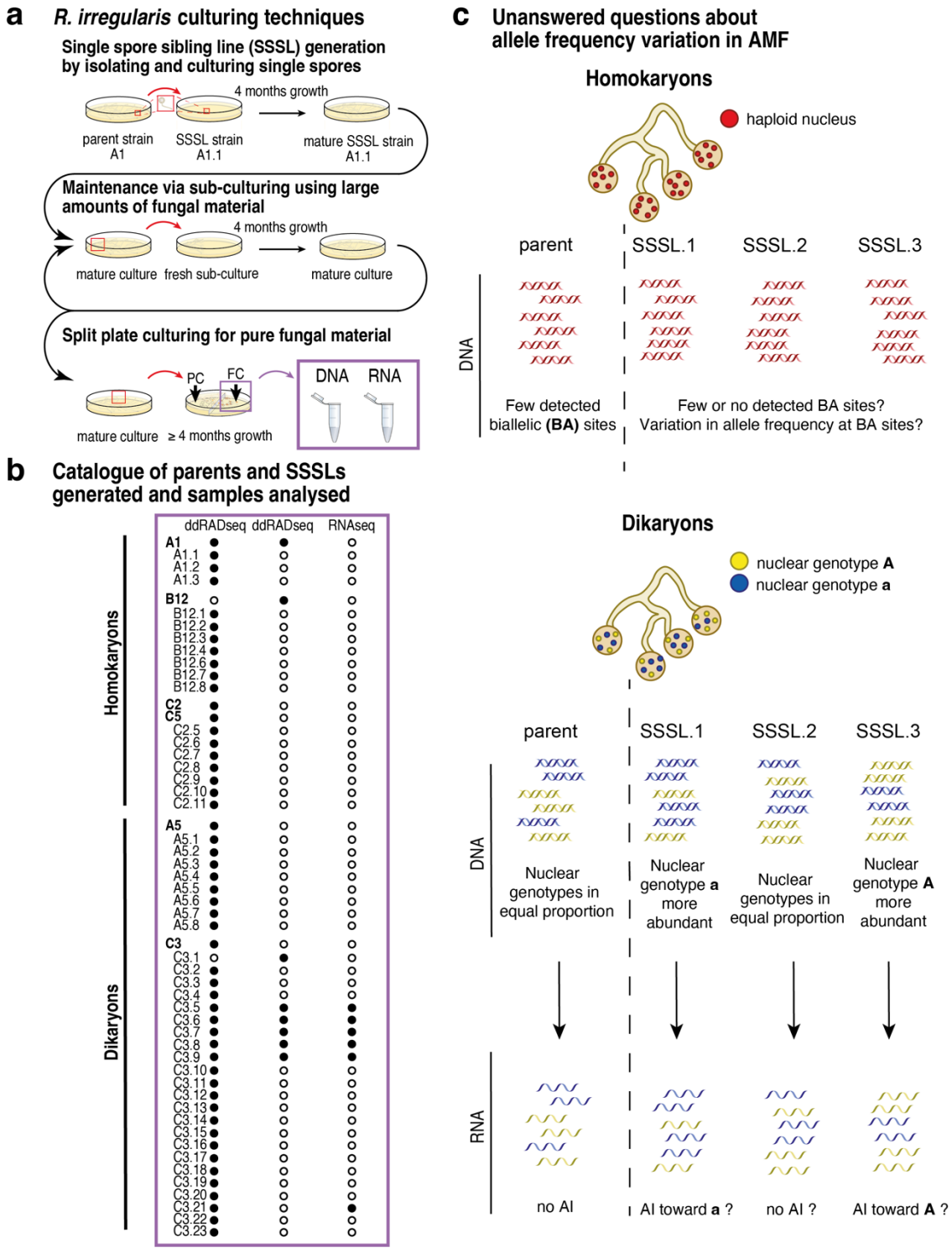


Figure 1. Experimental procedures, collected data and unanswered questions about genetic variation by clonally produced arbuscular mycorrhizal fungal (AMF) siblings. (**a**) Single spore culturing involves taking one spore to initiate a new culture and produce sibling cultures of a parental isolate. Sub-culturing, transferring a large amount of fungal material is used to maintain single spores sibling lines (SSSLs) and produce a larger amount of material for molecular analyses. PC refers to the plant compartment and FC refers to the fungal compartment. (**b**) Parental AMF lines and their SSSLs that were used for molecular analyses. Black filled dots indicate samples included in a given analysis, and white open dots indicate the samples were not used for a given analysis. **(c)** Schematic diagram of the unanswered questions posed in this study about generation of genetic variation among SSSLs and their gene transcription. Analysis at the genome level using DNA sequencing allows the test of whether siblings of a parental AMF isolate are genetically indistinguishable (as expected in homokaryon offspring) or genetically variable at the quantitative level (as was predicted in dikaryon offspring), while transcriptome analysis allows the test of whether allelic imbalance (AI) in gene transcription occurs in dikaryon siblings that display quantitative genetic variation at the genome level.

Materials and methods

Fungal material and growth conditions

R. irregularis isolates from Tänikon, Switzerland ('parental' isolates; homokaryons: A1, B12, C2; and dikaryons: A5 and C3) were used in this study (Koch et al., 2004). C5 also included in the analysis and is considered a clone of C2 as they are genetically indistinguishable (Savary et al., 2018a; Wyss et al., 2016). Forty-eight SSSLs were generated from these parents and maintained at 25°C in dark, axenic conditions with Ri T-DNA modified carrot roots (Fig. 1a) (Rosikiewicz et al., 2017; St-Arnaud et al., 1996). Additional cultures were produced independently for conducting a second ddRAD-seq (5 SSSLs of C3), as well as RNA-seq (6 SSSLs of C3), and were maintained in the same manner (Fig. 1b). Three individual split-plates (3 biological replicates; ddRAD-seq) or 3 pools each of 4 split-plates (3 biological replicates; RNA-seq) were produced.

DNA extraction, ddRAD-seq library preparations and sequencing

After at least 4 months, medium from fungal compartments containing AMF hyphae and spores was dissolved in 500 mL stirred citrate buffer (0.0062 M citric acid, 0.0028 M sodium citrate) for 20 min. One compartment represented one biological replicate (Fig. 1a). Fungal material was collected, flash-frozen and stored at -80°C until use.

Homogenized samples (CryoMill; Retsch GmbH, Haan, Germany) (2X 30 sec., 25 Hz, resting 30 sec., 5 Hz) were used to extract DNA (Qiagen Plant DNA kit; Qiagen, Hombrechtikon, Switzerland). DNA was quantified (Promega Quantus™ Fluorometer and DNA QuantiFluor® dye; Promega AG, Dübendorf, Switzerland) and stored at -20°C.

During ddRAD-seq library preparation, samples were subjected to duplicate digests to obtain two technical replicates of each sample (2 hrs at 37°C, then 20 min at 65°C: 1X CutSmart® buffer, 50 mM NaCl, 0.05 µg µL^{-1} BSA, 1 U *MseI*, 5 U *EcoRI-HF®*, 6µL template) using a frequent (*MseI*: New England Biolabs, Bioconcept AG, Allschwil, Switzerland) and a less frequent (*EcoRI-HF®*: New England Biolabs, Bioconcept AG, Allschwil, Switzerland) cutting restriction enzyme(Savary et al., 2018a; Wyss et al., 2016). DNA was diluted to 15 ng µL^{-1}, or used directly at lower concentrations. Adapters and barcodes were ligated (6hrs at 16°C, then 10 min at 65°C: 1X T4 ligase buffer, 14 mM NaCl, 0.014 µg µL^{-1} BSA, 862 nM *MseI* adapter, 86.3 nM *EcoRI* adapter, and 335 U T4 ligase; Table S1) and samples were purified (AMPure XP beads Beckman-Coulter; Indianapolis, United States; 1X bead volumes) before PCR. PCRs were performed in triplicate (30 sec. 98°C, 26 cycles (20 sec. 98°C, 30 sec. 60°C, 40 sec. 72°C), followed by 10 min. 72°C; 1X Q5® High Fidelity Buffer, 363 µM dNTPs, 305 nM forward and reverse primers (Table S1), 0.9X High GC Enhancer, and 0.4 U Q5® High Fidelity polymerase), verified by gel electrophoresis (1.5 % agarose gel, 100 V for 1 hr), size-selected (~300 bp; AMPure, 1X bead volume) and quantified before pooling. Equal quantities of ≤ 48 samples

were pooled per library (Table S2). Libraries were purified (AMPure, 1X bead volume) and verified (Fragment Analyzer) (Agilent, Santa Clara, United States) before sequencing.

The five SSSLs of C3 underwent the same procedure, but with doubled reaction volumes and were pooled and sequenced independently in a single library (Fig 1b).

Lausanne Genomic Technologies Facility sequenced 100 bp paired-end reads using Illumina® HiSeq 2500 (Illumina, San Diego, United States). Demultiplexed data files are deposited with European Nucleotide Archive under accession numbers PRJEB37069 (parental isolates and 48 SSSLs) and PRJEB39082 (5 SSSLs of C3).

RNA extraction, RNA-seq library preparation and sequencing

After 4 months, medium from fungal compartments of 6 SSSLs of C3 were dissolved in stirred citrate buffer for 50 minutes and washed with sterile ddH_2O. Four pooled compartments represented one biological replicate. Total RNA was extracted (Maxwell RSC Plant RNA kit, Promega) and RNA quantity and quality were determined (Nanodrop photometer and Aligent 5200 Fragment Analyzer). Two duplicate RNA-seq libraries were prepared using 100 ng RNA each and 13 cycles of PCR enrichment, representing technical replicates of each biological replicate (NEBNext Ultra II RNA Library Prep Kit for Illumina, NEB). Libraries with unique indices were pooled and 150 bp paired-end reads were sequenced with Illumina HiSeq 4000 platform in 3 lanes. Six replicates (2 technical replicates of 3 biological replicates) of each *R. irregularis* SSSL were sequenced. RNA-seq reads were deposited in the European Nucleotide Archive (PRJEB39188).

ddRAD-seq data pre-processing on 6 parental isolates and 48 SSSLs

Low-quality reads were removed using CASAVA filter (Y). Adapters and low-quality bases were trimmed using *tagcleaner.pl* (Schmieder et al., 2010). Only paired-reads with lengths >50 bp, and mean base quality >25, were retained with prinseq-lite-0.20.4 (Schmieder and Edwards, 2011). Demultiplexing was done with *process_radtags*, allowing ≤2 bp mismatch (Catchen et al., 2011). Sample reads were mapped to *R. irregularis* DAOM197198 (ASM43914v3) to assess qualitative differences. Quantitative analyses of allele frequencies at bi-allelic sites were achieved by mapping to the respective parental genome (PRJEB33553). Only uniquely mapped reads were considered using *bwa mem* algorithm with –c 2 (Table S2) (Li and Durbin, 2009). Variants with allele frequency ≥10% and coverage ≥20 were called using Freebayes 1.2.0 and only bi-allelic sites were considered further (Garrison and Marth, 2012). Variable sites were filtered (present in ≥60% of biological replicates) with *bcftools* to obtain common variants (Li et al., 2009). All scripts are available at: https://github.com/chanz06/AMF_RADseq_scripts.

Qualitative and quantitative analyses of ddRAD-seq data on parental isolates and 48 SSSLs

Samples containing ≥4000 SNPs were combined in a presence/absence matrix (165303 sites; missing information was considered as absent). These filters eliminated the parent isolate B12 from further

analyses. The *dendextend* 1.14.0 and *circulize* 0.4.10 R packages computed distances and generated a phylogenetic tree using the binary distance method. The package *gmodels* 2.18.1 was used to compute principal components using *fast.prcomp* function.

Common bi-allelic sites among a parent and all its SSSLs were selected and reads supporting the reference and alternative allele were used to compute allele frequencies. Allele frequencies at bi-allelic sites were quantitatively assessed using two methods. First, a traditional chi-squared test was used to detect significant differences between the reference allele abundance in the parent and SSSL at each bi-allelic site. Second, a non-parametric Mann-Whitney U test was used to test quantitative changes in allele frequencies between SSSLs and their parent. All bi-allelic sites and statistical testing results are documented in Tables S3a-d. Mean reference allele frequencies finally tested with *one-sample t-test* for significant shifts in SSSLs compared to their parent. All scripts are available at: https://github.com/chanz06/AMF_RADseq_scripts.

ddRAD-seq data pre-processing and mapping of 5 dikaryon SSSLs of C3

Adapter sequences were removed with *tagcleaner.pl* and low-quality reads were trimmed with *prinseq.pl* (Schmieder and Edwards, 2011; Schmieder et al., 2010). Only reads ≥50 bp were kept. Reads were aligned to *R. irregularis* A4 genome (PRJNA299206), using Novoalign V3.04.04 (Novocraft - Technologies, 2016). This assembly was used because previous analyses revealed high similarity to C3, and thus, these two isolates are considered genetically indistinguishable (Chen et al., 2018b; Savary et al., 2018a; Wyss et al., 2016). Mapping statistics can be found in Table S4.

The same exact methods were applied to diploid (*C. albicans* and *B. nana*) and tetraploid (*B. x intermedia*) controls. Publicly available ddRAD-seq data is retrievable at NCBI SRA database (BioProject Accession Numbers: PRJNA268659 and PRJEB3322) using the reference genomes GCA_000182965.3, GCA_000327005.1 (Wang et al., 2013).

RNA-seq data pre-processing and mapping of 6 dikaryon SSSLs of C3

Adapter sequences and low-quality bases were removed with Trimmomatic v0.36(Bolger et al., 2014). Reads were mapped onto the A4 genome (PRJNA299206) with STAR software v2.6.0, using the following parameters --alignIntronMin 20 --alignIntronMax 5000 --outFilterMismatchNoverLmax 0.4 --outFilterMismatchNmax 15 --sjdbOverhang 99 --outFilterIntronMotifs RemoveNoncanonical --alignEndsType EndToEnd --outSAMtype BAM SortedByCoordinate --outSAMattributes Standard --outSAMstrandField intronMotif(Dobin et al., 2013). Mapping statistics are contained in Table S5.

Variant calling, filtering and allele frequency estimations of bi-allelic sites of SSSLs of C3 from ddRAD-seq and RNA-seq data

Variant calling was performed in the same way for both datasets using Freebayes v1.2.0 (Garrison and Marth, 2012). SNPs, indels and multiple-nucleotide polymorphisms (MNPs) were detected with

coverage >10 with a diploid assumption (-p 2). Parameters -0 -J -K -u -F 0.1 ensured all possible variants were called.

SNPs in repeats were discarded using a repeat annotation file and *bedtools intersect* (Quinlan, 2014). Only bi-allelic sites with a Phred-scaled Qscore ≥30 and in scaffolds >1kb were considered, provided they were detected in all 6 replicates (Tables S6 and 7). Finally, only bi-allelic sites with a depth within the interquartile range of its sample, and ≥20 reads and in at least 5 replicates, were retained. We estimated the pooled reference allele frequency at all bi-allelic sites that did not vary significantly among replicates (chi-squared test, $p > 0.05$). Statistics and plots were performed using R software (Ihaka and Gentleman, 1996). A PCA of common bi-allelic sites in ddRAD-seq data was performed with the *prcomp* function. Plots were made with *ggplot2* package. Custom python scripts are available at https://github.com/jquimcrz/afreq_NGS.

Allelic fold-changes and allelic imbalance thresholds of genes in RNA-seq data

To investigate genes with evidence of disproportionate allelic transcription, $\log_2$ values of allelic fold-change (aFC) were estimated for each gene containing at least one bi-allelic site (passing filters). When a gene contained >1 bi-allelic position, the highest coverage position was used for estimates, meaning that aFC was based on allele frequencies at one bi-allelic site. Thresholds applied to determine allelic-imbalance were absolute values of $\log_2(aFC) > 0.5$. A *chi-squared* test compared the proportion of genes under allelic-imbalance among SSSLs using *prop.test*. A post-hoc pairwise comparison of proportions was performed with *pairwise.prop.test*, using Holm-Bonferroni corrections.

Functional annotation of genes under allelic-imbalance

We used eggNOG mapper to perform a functional annotation of identified genes under allelic-imbalance (Huerta-Cepas et al., 2019; Jensen et al., 2008). Results were summarized based on their Clusters of Orthologous Groups (COG) (Tatusov et al., 2003).

Identification of bi-allelic sites with mono-allelic and bi-allelic expression

The genomic bi-allelic sites were first identified by mapping whole genome sequencing reads of A4 (PRJNA299206) to the A4 assembly (Novoalign V3.04.04) and calling variants using the same parameters as described for 6 SSSLs (Garrison and Marth, 2012). Only bi-allelic sites in coding sequences of annotated genes, with ≥25 depth and ≥10 reads supporting both alleles, were further considered.

Using *samtools depth*, we then computed the number of reads mapping to bi-allelic positions (≥25 reads) from each RNA-seq alignment file and looked for the presence of either one (mono-allelic expression) or two (bi-allelic expression) alleles among mapped reads in these bi-allelic sites (Li et al., 2009).

FPKM analysis of genes with mono-allelic and bi-allelic expression

FPKM values were computed for genes using RSeQC "FPKM_count.py" (Wang et al., 2012). Significant differences ($p < 0.05$) between mono-allelic and bi-allelic expressed genes were determined by Mann-Whitney U test of transformed values ($\log_2$(FPKM+1)).

Results

Qualitative assessment of *R. irregularis* parental isolates and their 48 SSSLs

We detected 165303 polymorphic loci in the ddRAD-seq data from *R. irregularis* parental isolates and all 48 SSSLs. These polymorphic loci allowed us to infer qualitative similarity among the parents and their SSSLs. The relationship among parental isolates conformed to that expected from previous results (Masclaux et al., 2019; Savary et al., 2018a; Wyss et al., 2016). We recovered three distinct groups: a group with C3 and its offspring, a second group with C2, C5 and their offspring and a third group comprising A1, B12 and A5 and their offspring (Fig. 2a). Parents and their SSSLs showed a clear separation among three distinct clusters based on the presence and absence of multiple polymorphic sites (Fig. 2b). All SSSLs clustered similarly with their parent and were, therefore, considered qualitatively indistinguishable, as would be expected for offspring from clonal reproduction.

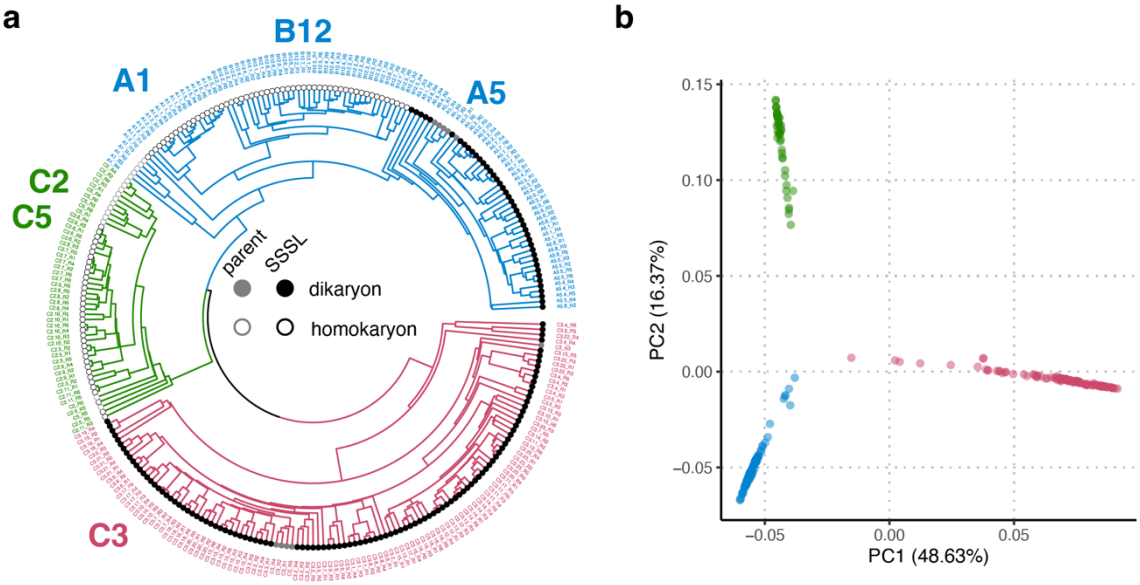


Figure 2. Qualitative analysis of 165303 bi-allelic sites in *R. irregularis* parental isolates and their 48 single spore sibling lines (SSSLs). (**a**) Relationship between parents and SSSL progeny based on shared polymorphisms. Homokaryon parents and SSSLs are labelled with grey and black open circles, respectively and dikaryon parents and SSSLs are labelled with grey and black filled circles, respectively. (**b**) Principal component analysis of present and absent polymorphisms showing distinct clustering of three groups. Colours follow the groupings shown in panel (a).

Quantitative assessment of bi-allelic sites in *R. irregularis* parental isolates and their 48 SSSLs

Very few bi-allelic sites were shared among homokaryon parents A1 and C2 and their SSSLs (167 and 32 sites, respectively) (Fig. 3a). In contrast, dikaryon isolates A5 and C3 shared more bi-allelic sites with their progeny (1233 and 299, respectively) (Fig. 3b). We tested whether this difference between dikaryons A5 and C3 was influenced by the high number of C3 SSSLs being compared. Indeed, we detected fewer common bi-allelic sites in dikaryons as we considered more SSSLs (Fig. 3c). There was a significant negative correlation between commonly detected bi-allelic sites and the

number of SSSLs compared (-0.6379, p = 0.0014 in homokaryons, and -0.8408, p < 0.001 in dikaryons).

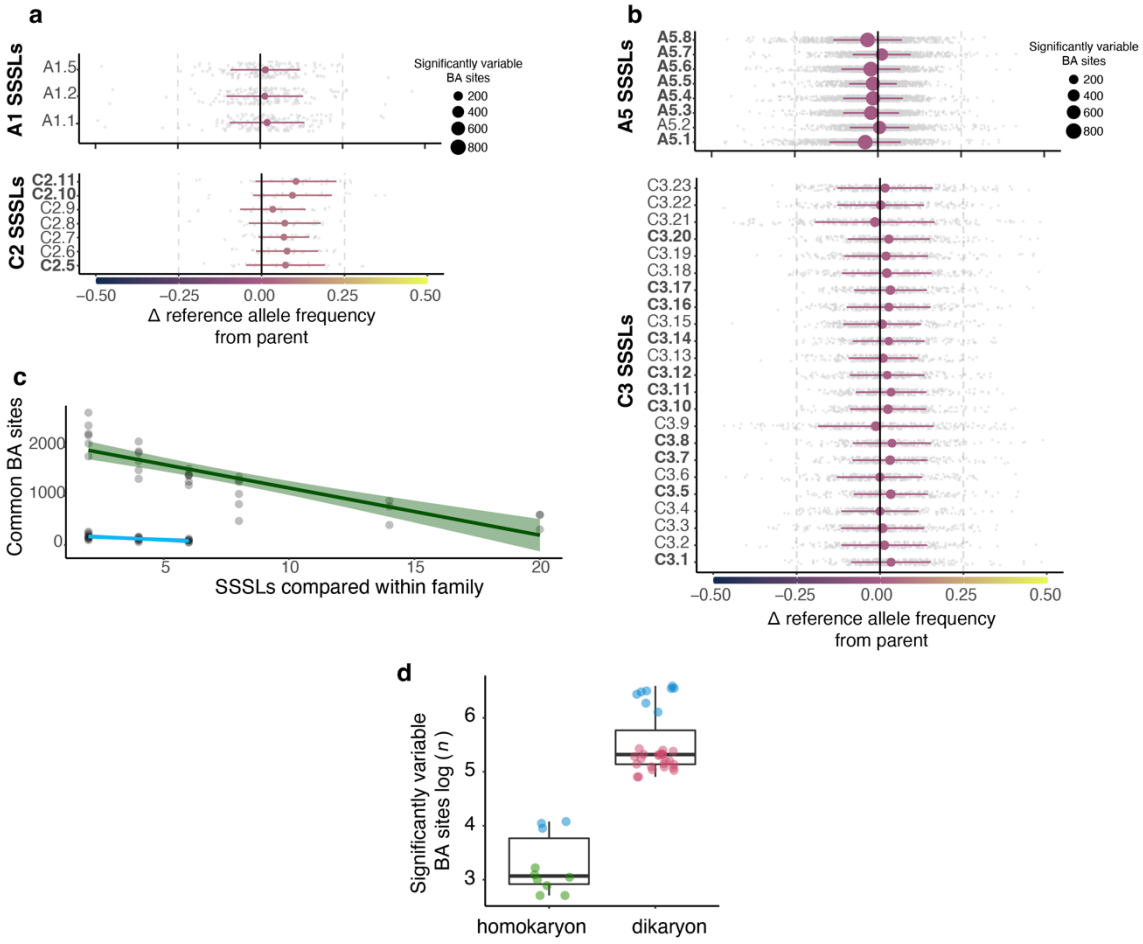


Figure 3. Quantitative changes in allele frequencies at bi-allelic (BA) sites between *R. irregularis* parental isolates and their single spore sibling lines (SSSLs). (**a**) Variation in reference allele (RA) frequencies among homokaryon SSSLs and their respective parent A1 and C2. The difference in RA frequency at each common BA site of a given sibling from its parent is plotted as a grey dot. The mean change in the RA frequency of an SSSL from its parent (and the standard deviation) are plotted with the point size indicating the number of significantly different BA sites (chi-squared, p < 0.05) The colour of the point indicates an increase or decrease in the RA compared to the parent. Significant overall shifts (t-test, p < 0.05) in SSSLs from parents are represented as bolded SSSL labels on the y axis. Significant shifts (t-test, p < 0.05) in allele frequencies of SSSLs are represented as bolded labels of SSSLs on the y axis. (**b**) Variation in RA frequencies between dikaryon SSSLs and their respective parent A5 and C3. Analysis and plots follow the same format as in (a). (**c**) Regression analysis of family (SSSLs originating from a parent) size effects on commonly detected BA sites. Homokaryons (light blue; *adjusted* R^2 = 0.3773, p = 0.0014) and dikaryons (green; *adjusted* R^2 = 0.6964, p < 0.0001). Green shading represents the 95% interval of confidence of the regression line. (**d**) Number of bi-allelic sites at which significant variation was found in allele frequencies in homokaryon and dikaryon *R. irregularis* families. Box-and-whisker plots show the distribution of significant BA sites for each SSSL. The scaled *log* values of the number of BA sites differing significantly in allele frequency between the parent and a SSSL (chi-squared, p < 0.05) is represented on the y-axis. The number of significantly variable loci between a parent and progeny, is significantly higher in dikaryons (t = - 13.088, df = 9, p = 3.665e-7; *one-sided t-test*). Green dots represent C2, C5 and SSSLs of C2. Blue dots in the homokaryon box represent A1 and its SSSLs. Red dots represent C3 and its SSSLs. The horizonal line represents the median, the box represents in the interquartile range and the vertical lines represent the maximum and minimum values.

We tested whether reference allele frequencies at common bi-allelic sites increased or decreased significantly (*i.e.,* quantitatively varied) in SSSLs relative to the frequencies in their parent. By subjecting read counts at bi-allelic sites to a chi-squared test, we found significant differences in relative allele frequencies between a parent and its offspring in both homokaryons and dikaryons (15-

59 sites and 135-727 sites, respectively) (Fig. 3a,b). The reference allele frequency differences between a parent and their SSSLs were significantly higher in dikaryons than in homokaryons (Fig. 3a,b,d).

We analysed mean reference allele frequencies at all significant bi-allelic sites between a parent and SSSL to understand whether these sites resulted in salient increases or decreases in reference allele frequencies of a given SSSL (Fig. 3a,b). We found that homokaryon SSSLs generally displayed fewer significant shifts in their reference allele frequency compared to their parent, and several experienced no significant change (Fig. 3a). Still, some differences were observed in some homokaryon SSSLs of C2, but may represent stochastic variation or positions at which there are potential problems in the genome assembly (Masclaux et al., 2019). More striking, were changes in reference allele frequency occurring in dikaryon SSSLs; notably, that changes were bi-directional, representing both reference allele increases and decreases in SSSLs compared to their parent (Fig. 3b). Most of A5 SSSLs retained a lower reference allele frequency compared to the parent (Fig. 3b). Only SSSL A5.7 showed a significant increase in the reference allele frequency compared to A5. There was a much broader range of variation among SSSLs of C3. The SSSLs C3.1, C3.5, C3.7, C3.8, C3.10, C3.11, C3.12 C3.14, C3.16, C3.17 and C3.20 all exhibited highly significant reference allele frequency increases compared to C3. The SSSLs C3.3, C3.6, C3.13, C3.15, C3.19, and C3.22 all showed similar reference allele frequencies to C3, and SSSLs C3.4, C3.9, and C3.21 showed decreases in the reference allele frequency compared to the parent C3.

Concordance of bi-allelic sites in ddRAD-seq and RNA-seq data from SSSLs of C3

We observed a large number of shared bi-allelic sites in the genome among SSSL replicates ranging from 1740-2318 in C3.6 and C3.8, respectively (Fig. S1). All SSSLs shared 1409 common genomic bi-allelic sites, with 684 located in coding regions (Fig. S2a,b). Of the 684 genomic bi-allelic sites observed in coding regions in ddRAD-seq data, only some of these were observed in the transcriptome, ranging from 130 to 144 in C3.6 and C3.21, respectively (Fig. S2c). RNA-seq reproducibility was lower than that observed in ddRAD-seq data, as many variants were unique to one technical replicate, being most likely sequencing artefacts due to the large differences in sequencing depth between the two experiments (Fig. S3 and Tables S4 and S5). Despite this, thousands of bi-allelic sites were consistent among replicates, and ranging from 5989 to 7117 in C3.8 and C3.21, respectively. Conservative posterior analyses of allele frequencies were restricted to a subset of these bi-allelic sites, resulting in from 479 to 757 (in C3.6 and C3.7, respectively; ddRAD-seq) and from 728 to 1445 bi-allelic sites (in C3.9, and C3.7, respectively; RNA-seq).

Variation in allele frequencies among 5 dikaryon SSSLs of C3

Allele frequency distributions of bi-allelic sites displayed the expected peaks the diploid (0.5) and tetraploid controls (0.25; 0.5; 0.75) (Fig. S4). Similarly, we examined allele frequencies among all SSSLs at common bi-allelic sites in ddRAD-seq data and revealed that SSSLs C3.5 (497 sites), C3.6

(479 sites) and C3.8 (739 sites) exhibited a unimodal allele frequency distribution centred at 0.5 (Fig. 4a). In contrast, two other SSSLs, C3.7 (757 sites) and C3.9 (604 sites), displayed bimodal distributions with peaks around 0.45 and 0.55 in C3.7; and 0.40 and 0.60 in C3.9 (Fig. 4a). Furthermore, even though both of these SSSLs displayed bimodal distributions, reference allele frequencies were opposing. More specifically, at a given site, the reference allele frequency was higher in C3.9 and the alternative allele frequency was higher in C3.7 (Fig. 4b). Using allele frequencies at 125 common bi-allelic sites, PCA revealed SSSL dissimilarity at these bi-allelic sites, explaining 63.3% (PC1) of the variance (Fig. 4c). We observed that SSSLs were distributed along PC1, likely representing the variation in abundance of nuclear genotypes among SSSLs. The three dikaryon SSSLs (C3.5, C3.6, and C3.8) that showed a 1:1 ratio of both nuclei (unimodal distributions) clustered together at the centre. The two dikaryon strains (C3.7 and C3.9) that displayed unequal proportions (bimodal distributions) of allele frequencies were diametrically opposed along PC1 (Fig. 4c).

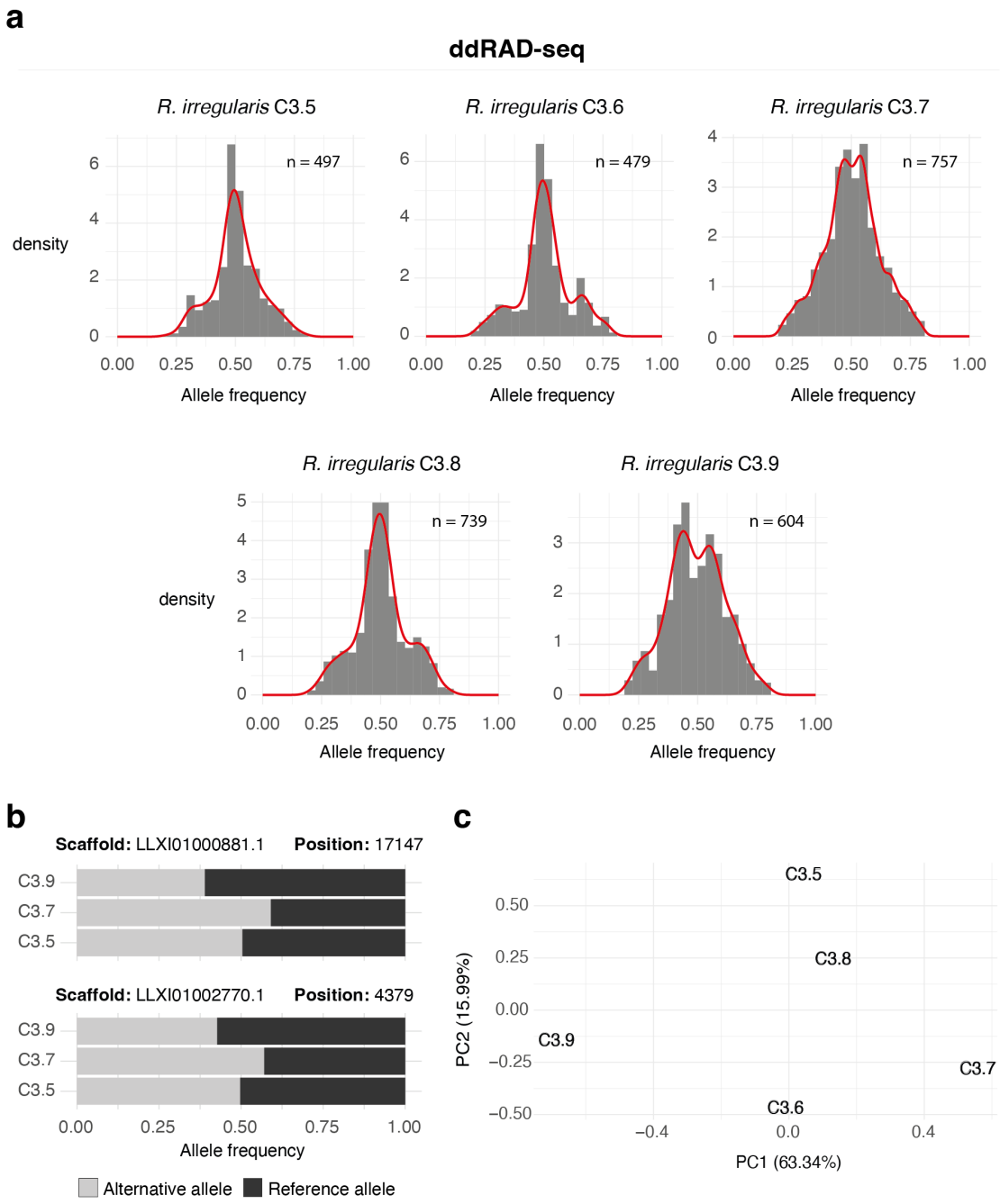


Figure 4. Analysis of ddRAD-seq data of 5 dikaryon single spore sibling lines (SSSLs) of C3. (**a**) Distribution of allele frequencies of bi-allelic (BA) sites in *R. irregularis* SSSLs. (**b**) Two examples of BA sites showing differences in their relative proportions between the two siblings C3.7 and C3.9. (**c**) Principal component analysis (PCA) of allele frequencies at BA sites. Score plot based on the allele frequencies of common BA sites of the SSSLs (n = 577 sites). The first two principal components are shown with their respective percentages of explained variance.

Transcriptome-wide differences in allele expression among 6 dikaryon SSSLs of C3

We then addressed whether RNA-seq data revealed transcriptional bias at bi-allelic sites in dikaryon SSSLs. A remarkably similar allele frequency distribution to that observed in the genomic data also occurred in SSSL transcriptomes (Fig. 5a). Similar to ddRAD-seq, bi-allelic sites in SSSLs C3.5 (1276 sites), C3.6 (1152 sites) and C3.8 (913 sites) showed unimodal distributions centred at 0.5. SSSLs C3.7 (1445 sites) and C3.9 (728 sites) again presented clear bimodal distributions in their transcript frequencies, similar to unequal allele frequencies observed in ddRAD-seq. C3.21 (1368 sites) also exhibited a bimodal distribution with the most extreme allele frequencies transcribed of all SSSLs (an approximate 3:7 ratio).

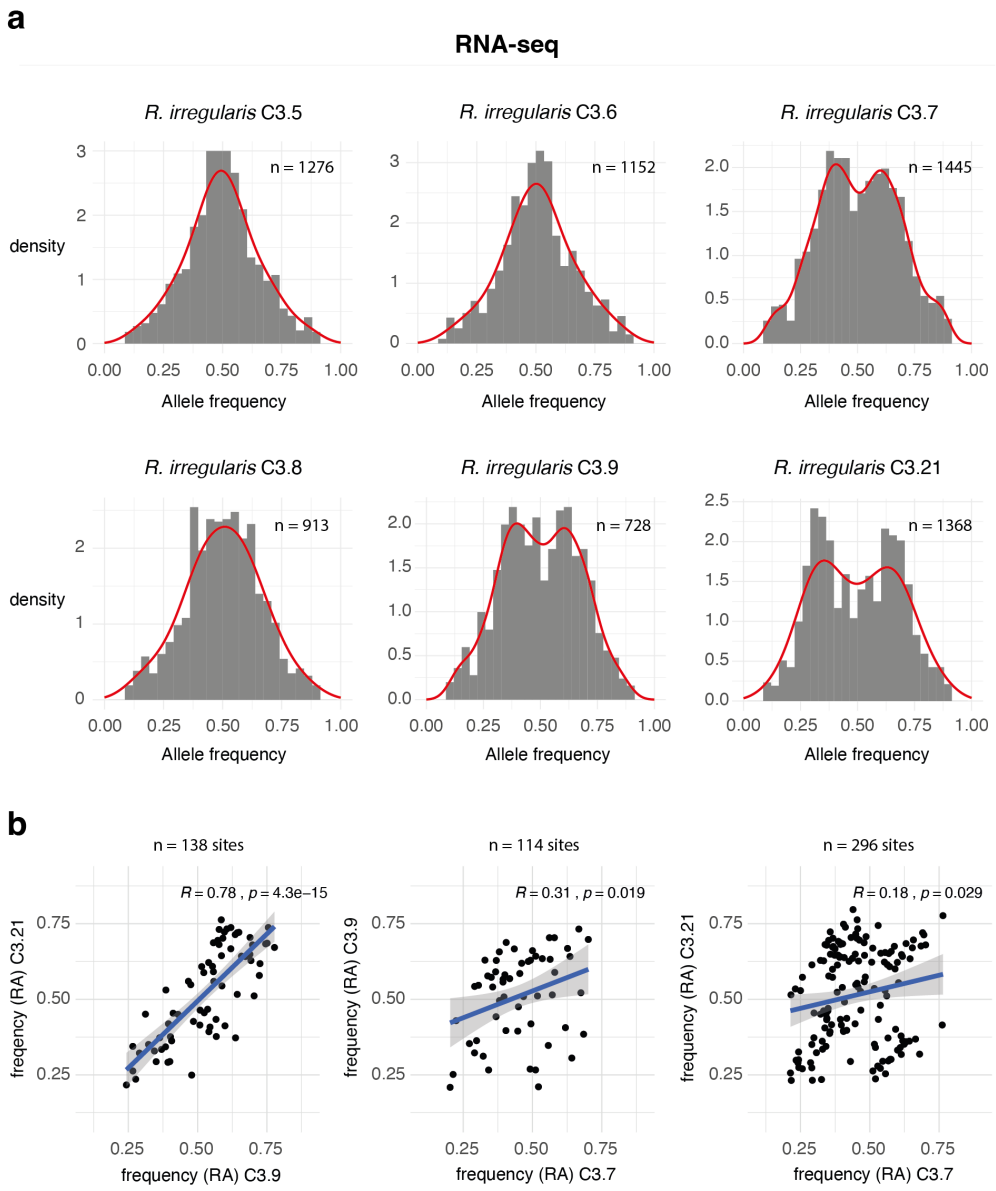


Figure 5. Analysis of allele frequencies on RNA-seq data. **(a)** Distribution of allele frequencies of bi-allelic (BA) sites of 6 dikaryon single spore sibling lines (SSSLs) of C3. (**b**) Pairwise comparisons of allele frequencies of common BA sites between *R. irregularis* strains C3.7, C3.9 and C3.21. Correlation of the allele frequencies (reference allele) at common BA sites (Pearson's correlation coefficient and probability are shown). Grey shading represents the 95% interval of confidence of the regression line.

We compared transcript allele frequencies at common bi-allelic sites to determine similarity among SSSLs with bimodal distributions. Pairwise comparisons between C3.9 and C3.21 allele frequencies revealed a positive correlation (R = 0.78) and a transcription bias towards the same, most abundant allele (Fig. 5b). In contrast, pairwise comparison of C3.7 with C3.9 (R = 0.31) and C3.7 with C3.21 (R = 0.18) showed much weaker correlation. This result is congruent with observations of ddRAD-seq data, where reference allele frequencies of SSSLs C3.7 and C3.9 were opposing.

Genes under allelic-imbalance during transcription

Similar to ddRAD-seq, and global RNA-seq analyses, we further observed bimodal distributions in SSSLs C3.7, C3.9 and C3.21 when testing for allelic fold-change variation in gene transcripts based on one bi-allelic site (Fig. 6a). Genes exhibiting allelic-imbalance were present in all 6 SSSLs, even though allele frequency distributions centred at 0.5 (Table S8). Still, allelic-imbalance of bi-allelic expressed genes significantly differed among the SSSLs (χ^2 statistic = 152.71, df = 5, p < 2.2e-16), and indeed, was more pronounced in SSSLs with bimodal allele frequency distributions. For example, C3.21 showed the highest proportion of genes under allelic-imbalance (close to 80%) and a slightly lower proportion in C3.7 and C3.9 (60-70%) (Fig. 6b; Table S9). Allele frequencies of several genes differed by up to ~25% between SSSLs C3.7 and C3.21 and were consistently dissonant (Fig. 6c, top). Other bi-allelic expressed genes exhibited similar allele frequencies among the SSSLs (Fig. 6c, bottom). Notably, most genes under allelic-imbalance were unique to individual SSSLs (Fig. S5a) and possessed a wide variety of biological functions (Fig. S5b).

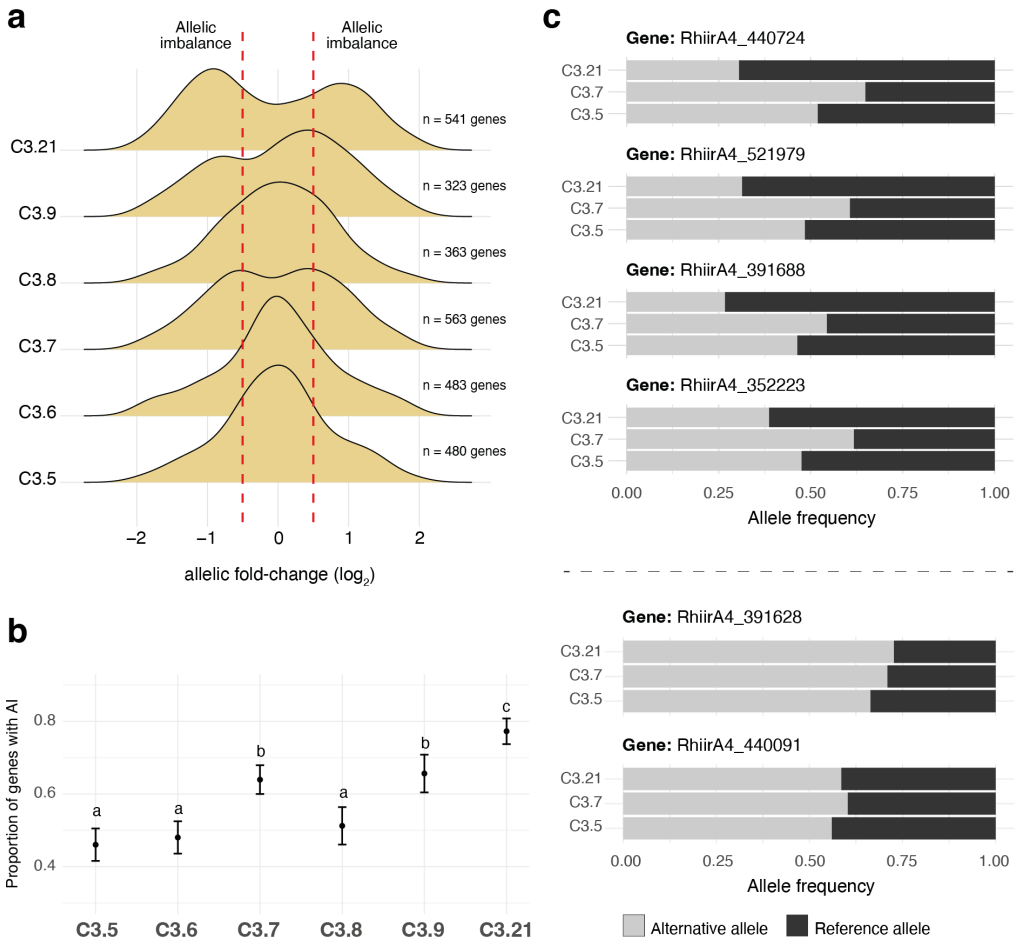


Figure 5. Analysis of allele frequencies on RNA-seq data. **(a)** Distribution of allele frequencies of bi-allelic (BA) sites of 6 dikaryon single spore sibling lines (SSSLs) of C3. (**b**) Pairwise comparisons of allele frequencies of common BA sites between *R. irregularis* strains C3.7, C3.9 and C3.21. Correlation of the allele frequencies (reference allele) at common BA sites (Pearson's correlation coefficient and probability are shown). Grey shading represents the 95% interval of confidence of the regression line.

Mono-allelic expression of bi-allelic sites within genes

RNA-seq data at bi-allelic sites revealed that either both alleles, or sometimes only one allele, was transcribed (Fig. 7a,b). We, therefore, further investigated the prevalence of mono-allelic expression at bi-allelic sites, and found that the number of bi-allelic sites with bi-allelic expression was lower than those with mono-allelic expression (Fig. 7c). Still, bi-allelic expressed genes were significantly higher than mono-allelic expressed genes in all SSSLs (Fig. 7d). Approximately 600 bi-allelic expressed and 250 mono-allelic expressed genes were identified in each SSSL, of which, 459 and 187 were commonly shared among all six SSSLs (Fig. 7e,f). Most notably, mono-allelic expressed genes had significantly higher SNP densities, compared to bi-allelic expressed genes (Fig. 7g) and were significantly less expressed than bi-allelic expressed genes (Fig. S6). In both cases, the functional annotation of genes with mono-allelic expression and bi-allelic expression revealed orthologs involved in many, and sometimes common, biological processes, such as energy production and conversion, transcription or signal transduction mechanisms (Tables S10 and S11).

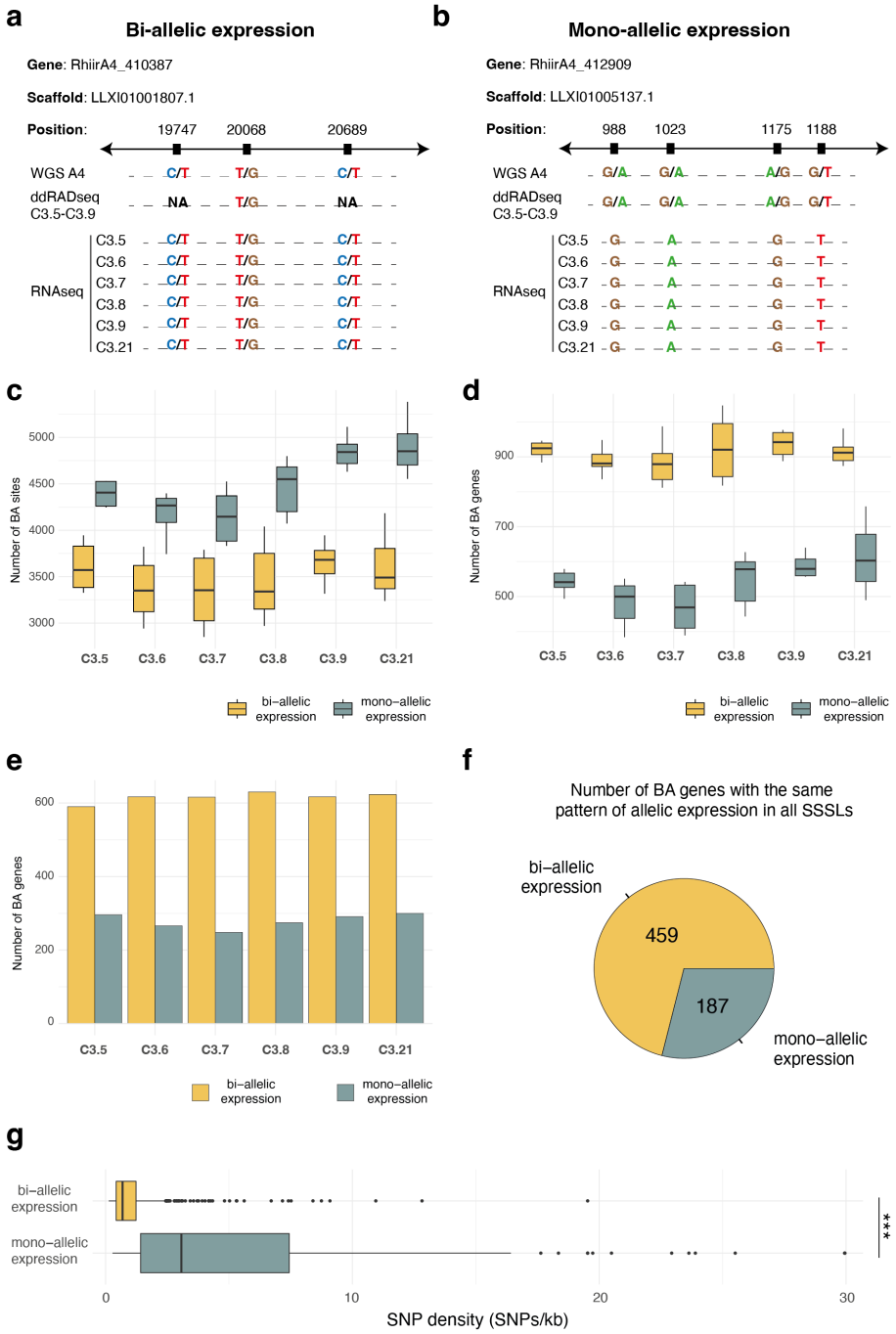


Figure 7. Expression of bi-allelic (BA) sites in the transcriptomes of 6 single spore sibling lines (SSSLs) of C3. (**a**) Example of a BA gene that exhibited BA expression in all SSSLs. (**b**) Example of a BA gene that exhibited mono-allelic (MA) expression of its BA sites in all SSSLs. (**c**) Number of BA sites where both alleles were expressed (BA expression) and number of BA sites where only one allele was expressed (MA expression). The horizonal line represents the median, the box represents in the interquartile range and the vertical lines represent the maximum and minimum values. (**d**) Number of BA genes that show BA expression or MA expression. The horizonal line represents the median, the box represents in the interquartile range and the vertical lines represent the maximum and minimum values. (**e**) Number of BA genes with BA expression or MA expression that were consistent among all replicates of each SSSLs. **(f)** Number of BA genes that showed BA expression or MA expression consistently among all replicates of all SSSLs. (**g**) Density of polymorphic sites (SNPs kb^{-1}) in genes that exhibited BA expression and MA expression. The horizonal line represents the median, the box represents in the interquartile range and the vertical lines represent the maximum and minimum values.

Discussion

In this study, we generated ddRAD-seq data on a cohort of 48 homokaryon and dikaryon SSSLs of their *R. irregularis* parental isolates. We showed that SSSLs are indeed clonal offspring, but that dikaryon SSSLs, despite qualitatively being clones, commonly exhibit quantitative allele frequency variation at bi-allelic sites. This variation represents proportions of two genetically distinct nuclei. Analysis on a subsample of dikaryon SSSLs from one parent revealed that the frequency of two nuclear genotypes deviated considerably from the parent. Ultimately, this translated into the predominance of one of the two nuclear genotypes in some SSSLs. Both nuclear genotypes contributed to gene transcription and the transcription of biallelic genes mirrored nuclear genotype frequencies. MA expression also sometimes occurred in genes that were BA and this was was more likely if with a greater divergence between alleles (*i.e,.* a higher SNP kb $^{-1}$ density) of the gene.

R. irregularis dikaryons produce clonal SSSLs which quantitatively differ in nuclear genotype proportions

Using ddRAD-seq data, we analysed more SSSLs than previous studies, and many more than would be possible with single nuclei sequencing. Multiple loci enabled us to assess genetic variation in homokaryon and dikaryon SSSLs (Fig. 8a). SSSLs clustered with their parents, indicating no significantly detectable qualitative genetic variation. Although a small number of bi-allelic sites were still detected in homokaryons, we expect that they have little to no functional consequence (Masclaux et al., 2019). Although the lack of genetic variation among homokaryon SSSLs is intuitive, it is interesting in the context of recent field-based experiments. Large significant differences in cassava yield were observed in the field in a fully replicated randomized-block deign experiment when cassava was inoculated with SSSLs originating from homokaryon parents (Ceballos et al., 2019; Peña et al., 2020). The SSSLs were the same ones on which ddRAD-seq was performed in this study. It is, therefore, improbable that yield differences induced by inoculation with different SSSLs can be attributed to quantitative genetic variation among homokaryon SSSLs, and so, likely depends on additional, contextual factors including potential epigenetic differences among SSSLs and how SSSLs affect soil microbial community composition and succession (Gao et al., 2019).

In contrast to homokaryons, we observed much more quantitative variation among dikaryon SSSLs. As hypothesized, bi-allelic sites were more prolific among dikaryon SSSLs, and reflect the presence of two genetically distinct nuclei. Furthermore, reference allele frequencies at multiple bi-allelic sites quantitatively deviated between clonal SSSLs and their parent, indicating the inheritance of different nuclear genotype proportions.

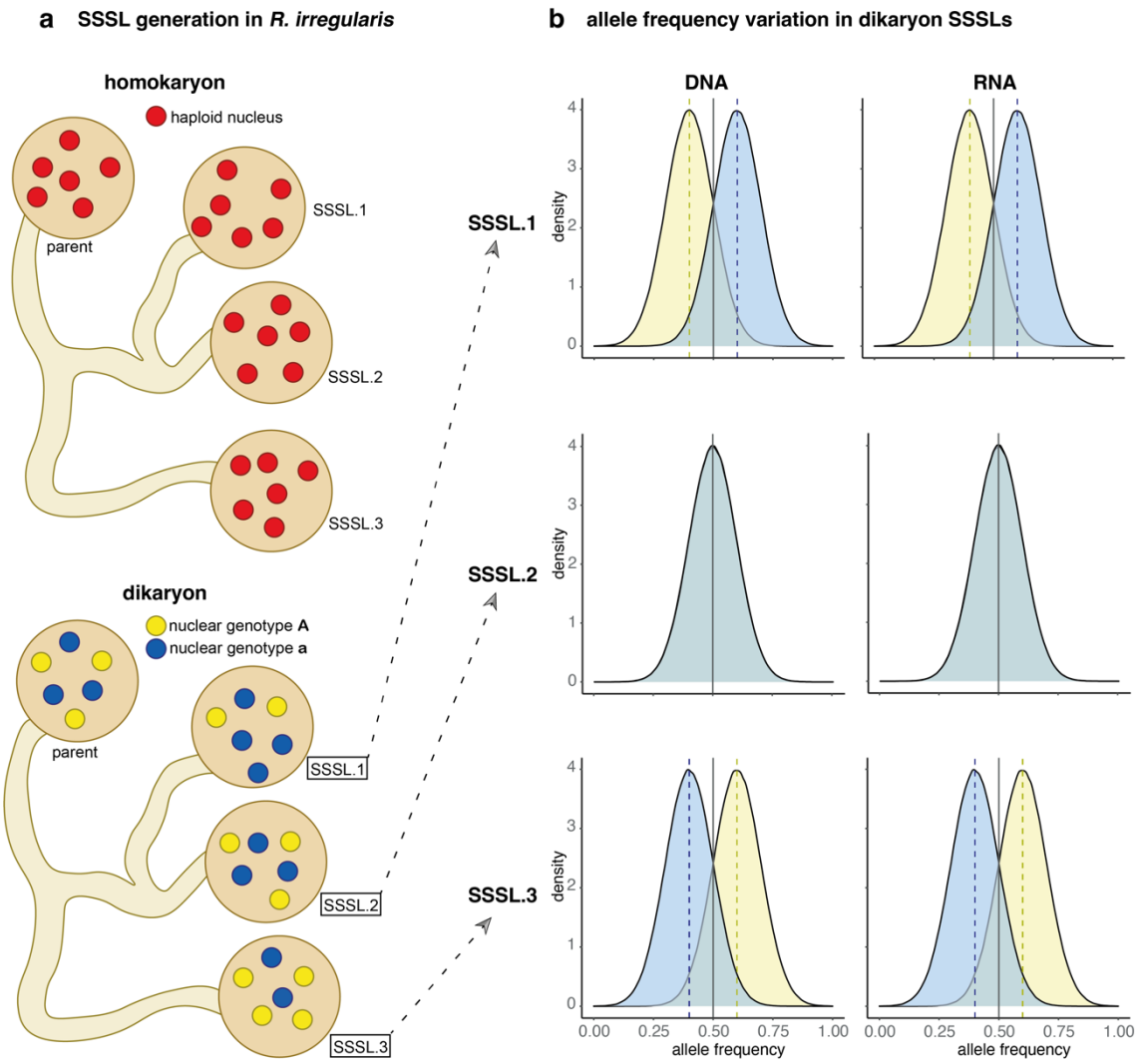


Figure 8. A schematic diagram summarizing the relationship between observed proportions of nuclear genotypes in *R. irregularis* parents and offspring and their gene transcription patterns. **(a)** A parental homokaryon isolate gives rise to clonally identical offspring (shown as red nuclei). A dikaryon parental isolate gives rise to offspring with relative proportions of two nuclear genotypes that can diverge in a single spore sibling line (SSSL) from the parental isolate (e.g. SSSL 1 and SSSL 3). Nuclei of the two different genotypes shown in yellow and blue. These proportions represent allele frequencies estimations in ddRAD-seq data. **(b)** Allele frequencies at bi-allelic sites in the genome and in the transcriptome of the three SSSLs 1, 2 and 3 shown in (a). Colours under the allele frequency curves represent alleles originating from each of the two nuclear genotypes.

Typical allele frequency distributions at bi-allelic sites in diploid organisms are unimodally distributed and centred at 0.5 (Zhu et al., 2016) (Fig S4). Similarly, AMF isolates with a population of two distinct haploid nuclear genotypes (*e.g.,* a dikaryon) should display diploid-like allele frequency distributions (Ropars et al., 2016). On the other hand, disproportionate inheritance of nuclei would result in deviations from 0.5 (Masclaux et al., 2018). We provide additional support based on more detailed analyses that SSSLs of C3 varied between 2:3 and 3:2 in nuclear ratios from their parent (1:1; alternative:reference allele frequency). Specifically, our results strongly indicate disproportionate inheritance of two nuclear genotypes in C3.7 and C3.9 and that quantitative genetic variation often occurs among dikaryon SSSLs (Fig. 8a). A previous, single locus study of *bg112* allele frequencies arrived at a similar conclusion, but, despite adequate replication, was scrutinized due to possible PCR

variability (Kokkoris et al., 2020; Masclaux et al., 2018). We confirm earlier results, and can conclude that this criticism is highly unlikely, given that independent and well replicated datasets produced near identical results across hundreds of bi-allelic sites. An alternative explanation for quantitative genetic differences observed among siblings would be that nuclei fused and recombined. However, this is unlikely. Single nucleus sequencing of *R. irregularis* isolate A4 (which is genetically indistinguishable from C3 and is, thus, considered a clone) revealed no evidence (Chen *et al*., 2018a) of among-nucleus recombination in this fungus. Although the same study detected a very small amount of recombination among nuclei of another isolate of the same species, this remains controversial (Auxier and Bazzicalupo, 2019).

Allelic imbalance in gene transcription in *R. irregularis* dikaryons

Previously, researchers suggested that gene expression in a dikaryon isolate might reflect proportions of both nuclei (Masclaux et al., 2018). Because some SSSLs of C3 displayed different nuclear genotype proportions, we wanted to see if evidence of the same could be found in transcriptome profiles of these SSSLs. The results confirmed that SSSLs with disproportionate nuclear genotypes based on ddRAD-seq data, also displayed allelic-imbalance in bi-allelic transcripts reflecting allele frequencies found in ddRAD-seq (Fig. 8b). These observations suggest a direct consequence of unequal nuclear genotype ratios on transcribed alleles, with the most abundant genotype being transcriptionally overrepresented. This is an important result because this indicates that the generation of such quantitative genetic variation could also potentially influence the AMF phenotype. Furthermore, because a previous study has shown associations between patterns of genome variation in *R. irregularis* and plant growth, such alterations in nuclear genotype frequency could potentially influence the symbiosis with plants (Ceballos et al., 2019).

Exceptions to the rule: When transcribed alleles do not reflect nuclear genotype ratios

Allele frequencies of transcripts did not always reflect the estimated nuclear genotype ratios, but this represented a much smaller number of bi-allelic genes than those that were expressed in the same proportion as the nuclear genotype frequencies. Interestingly, some bi-allelic genes exhibited the same pattern of allelic-imbalance in expression in all SSSLs, regardless of nuclei proportions (Fig. 6c). Therefore, it is likely that some genes are affected by other transcriptional regulatory mechanisms that are independent of nuclear genotype proportions.

Bi-allelic versus Mono-allelic expression at bi-allelic sites suggests multi-layered regulation of transcription in dikaryons

We observed that at many bi-allelic sites in the genome, only one of the two possible sequence variants was actually transcribed. This was a consistent and significant pattern across all six SSSLs irrespective of nuclear genotype ratios (Fig. 7). Intriguingly, significantly more bi-allelic genes expressed both alleles, rather than showing mono-allelic expression. Again, this was a remarkably similar pattern across all SSSLs and across replicates of each SSSL, revealing a very robust pattern (Fig. 7). Taken together, these results show that bi-allelic genes in which greater divergence between the two alleles has occurred (as measured by the number of bi-allelic sites in the gene) are less likely to both be transcribed. These results point both to biased mono-allelic expression at highly polymorphic sites, as well possible epigenetic silencing of highly divergent alleles; a situation that is predicted in a conflictual scenario between two divergent genomes (Dyson and Goodisman, 2020; Zou et al., 2020).

One other completely unexpected result was that transcription was also consistently significantly higher in bi-allelic expressed genes compared to mono-allelic expressed genes across all replicates and all SSSLs (Fig. S6). The fact that transcription of both alleles of a gene gave rise to significantly more transcripts than those exhibiting mono-allelic expression, suggests that in more highly divergent bi-allelic genes, suppression of one allele limits the transcription of the gene. A prediction from this finding would be that genes that are required to respond to a sudden environmental cue by rapidly producing a high transcript number should be under selection to retain two alleles that have undergone little divergence. However, we cannot completely exclude the possibility that overall lower expression levels of mono-allelic expressed genes hindered the detection of the second allele in some cases.

It was important that all cultures were maintained in a homogeneous environment as to not influence transcription results. In cases where mono-allelic expression occurs in a bi-allelic gene, it is also possible that selection would favour the retention of two divergent alleles that could be differentially expressed in different environments. Experimentally manipulated environments may shed light on this possibility.

Ecological significance and application of quantitative variation among dikaryon SSSLs

Fungi typically display an array of nuclear dynamics to fit their life strategies. For example, yeasts, which are not host dependent, show a fitness cost associated with being diploid, and thereby, haploid strains adapt and evolve much faster (Marad et al., 2018). On the other hand, the obligate plant pathogenic rust fungi, *Puccinia graminis* f. sp. *trici*, needs two plant hosts to complete its lifecycle, but can only infect alternate hosts with homokaryon spores, and primary hosts with dikaryon spores (Bakkeren and Szabo, 2020). Scott et al. (2019) recently compared two models of AMF evolution in which selection acts either on individuals, or on the nucleus. The current opinion is that high intraspecific genetic diversity in *R. irregularis* could facilitate evolution by enabling generalist lifestyles, and overcoming the danger of becoming too specialized on one host (Chen et al., 2018b). This relationship was explored and demonstrates that nuclear dynamics may change in response to

particular plant hosts for dikaryon AMF (Kokkoris et al., 2021; Angelard et al., 2014). In nature, AMF dikaryons may optimize niche adaptation in multiple ecosystems by maintaining both populations of cooperating nuclear genotypes. Dikaryon SSSLs indeed exhibit large differences in quantitative traits and affect plant growth significantly (Angelard et al., 2010; Ceballos et al., 2019; Ceballos et al., 2013; Peña et al., 2020). This might perhaps be due to the fact that SSSLs with the most optimal ratios of nuclear genotypes colonise and form symbioses with a given host more rapidly.

In conclusion, we show that dikaryon *R. irregularis* isolates commonly generate quantitative shifts in allele frequencies among single spore offspring. These shifts in allele frequencies are observed in hundreds of bi-allelic sites across the genome and likely reflect the changes in proportions of the two nuclear genotypes. We further conclude that varying nuclear dynamics of SSSLs generate similar quantitative shifts in gene transcription, meaning that transcription is linked to the underlying nuclear ratios of SSSLs. These findings hint toward additional factors to consider that may regulate transcription and symbiosis within these important plant mutualists.

Data availability

All sequencing data are deposited in European Nucleotide Archive under the following accessions: PRJEB37069, PRJEB39082, PRJEB39188.

Acknowledgements

The authors would like to thank Jérémy Bonvin and Paweł Rosikiewicz for previously generating SSSLs used in this study. We would also like to thank Tania Wyss for generating the ddRAD-seq dataset for the 5 SSSLs of C3. Finally, we appreciate the staff at the Lausanne Genomics Technology Facility (GTF) for sequencing DNA and RNA libraries. The research was funded by the Swiss National Science Foundation (Project number: 310030B_182826).

Author contribution

CR, JCC, and IRS designed experiments; CR, CA, and RS produced ddRAD-seq data on 48 SSSLs; JCC, CA, SJL and RS produced RNA-seq data on 6 dikaryon SSSLs; FM generated ddRAD-seq on 5 dikaryon SSSLs; CR and JCC analysed sequencing data; CR, JCC, SJL, IDM and IRS interpreted data and wrote the manuscript; IRS acquired project funding.

References

Angelard C, Colard A, Niculita-Hirzel H, et al. (2010) Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Current Biology* 20(13): 1216-1221.

Angelard C, Tanner CJ, Fontanillas P, et al. (2014) Rapid genotypic change and plasticity in arbuscular mycorrhizal fungi is caused by a host shift and enhanced by segregation. *ISME Journal* 8(2): 284-294.

Bago B, Pfeffer PE and Shachar-Hill Y (2000) Carbon metabolism and transport in arbuscular mycorrhizas. *Plant Physiology* 124(3): 949-958.

Bakkeren G and Szabo LJ (2020) Progress on Molecular Genetics and Manipulation of Rust Fungi. *Phytopathology* 110(3): 532-543.

Bolger AM, Lohse M and Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114-2120.

Bravo A, Brands M, Wewer V, et al. (2017) Arbuscular mycorrhiza-specific enzymes FatM and RAM2 fine-tune lipid biosynthesis to promote development of arbuscular mycorrhiza. *New Phytologist* 214(4): 1631-1645.

Brundrett MC and Tedersoo L (2018) Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist* 220(4): 1108-1115.

Catchen JM, Amores A, Hohenlohe P, et al. (2011) Stacks: building and genotyping Loci de novo from short-read sequences. *G3* 1(3): 171-182.

Ceballos I, Mateus-Gonzalez ID, Peña R, et al. (2019) Using variation in arbuscular mycorrhizal fungi to drive the productivity of the food security crop cassava. *bioRxiv*. DOI: https://doi.org/10.1101/830547.

Ceballos I, Ruiz M, Fernandez 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. *Plos One* 8(8): e70633.

Chen ECH, Mathieu S, Hoffrichter A, et al. (2018a) Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. *Elife* 7.

Chen ECH, Morin E, Beaudet D, et al. (2018b) High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytologist* 220(4): 1161-1171.

Clergeot PH, Rode NO, Glemin S, et al. (2019) Estimating the Fitness Effect of Deleterious Mutations During the Two Phases of the Life Cycle: A New Method Applied to the Root-Rot Fungus Heterobasidion parviporum. *Genetics* 211(3): 963-976.

Croll D, Giovannetti M, Koch AM, et al. (2009) Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus Glomus intraradices. *New Phytologist* 181(4): 924-937.

Davison J, Moora M, Opik M, et al. (2015) Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* 349(6251): 970-973.

Dobin A, Davis CA, Schlesinger F, et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1): 15-21.

Dong X, Zhang M, Chen J, et al. (2017) Dynamic and Antagonistic Allele-Specific Epigenetic Modifications Controlling the Expression of Imprinted Genes in Maize Endosperm. *Molecular Plant* 10(3): 442-455.

Dyson CJ and Goodisman MAD (2020) Gene Duplication in the Honeybee: Patterns of DNA Methylation, Gene Expression, and Genomic Environment. *Mol Biol Evol* 37(8): 2322-2331.

Ehinger MO, Croll D, Koch AM, et al. (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(3): 853-861.

Fellbaum CR, Gachomo EW, Beesetty Y, et al. (2012) Carbon availability triggers fungal nitrogen uptake and transport in arbuscular mycorrhizal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* 109(7): 2666-2671.

Gao C, Montoya L, Xu L, et al. (2019) Strong succession in arbuscular mycorrhizal fungal communities. *ISME Journal* 13(1): 214-226.

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

Gehrmann T, Pelkmans JF, Ohm RA, et al. (2018) Nucleus-specific expression in the multinuclear mushroom-forming fungus Agaricus bisporus reveals different nuclear regulatory programs. *Proc Natl Acad Sci U S A* 115(17): 4429-4434.

Govindarajulu M, Pfeffer PE, Jin HR, et al. (2005) Nitrogen transfer in the arbuscular mycorrhizal symbiosis. *Nature* 435(7043): 819-823.

Huerta-Cepas J, Szklarczyk D, Heller D, et al. (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47(D1): D309-D314.

Ihaka R and Gentleman R (1996) R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* 5(3): 299-314.

Jensen LJ, Julien P, Kuhn M, et al. (2008) eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res* 36(Database issue): D250-254.

Keymer A, Pimprikar P, Wewer V, et al. (2017) Lipid transfer from plants to arbuscular mycorrhiza fungi. *Elife* 6: e29107.

Koch AM, Kuhn G, Fontanillas P, et al. (2004) High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. *Proc Natl Acad Sci U S A* 101(8): 2369-2374.

Kokkoris V, Chagnon PL, Yildirir G, et al. (2021) Host identity influences nuclear dynamics in arbuscular mycorrhizal fungi. *Current Biology*. Epub ahead of print 2021/02/06. DOI: 10.1016/j.cub.2021.01.035.

Kokkoris V, Stefani F, Dalpe Y, et al. (2020) Nuclear Dynamics in the Arbuscular Mycorrhizal Fungi. *Trends in Plant Science* 25(8): 765-778.

Lafon-Placette C, Hatorangan MR, Steige KA, et al. (2018) Paternally expressed imprinted genes associate with hybridization barriers in Capsella. *Nature Plants* 4(6): 352-357.

Larsson AJM, Johnsson P, Hagemann-Jensen M, et al. (2019) Genomic encoding of transcriptional burst kinetics. *Nature* 565(7738): 251-254.

Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14): 1754-1760.

Li H, Handsaker B, Wysoker A, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 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(1): e1004078.

Marad DA, Buskirk SW and Lang GI (2018) Altered access to beneficial mutations slows adaptation and biases fixed mutations in diploids. *Nature Ecology & Evolution* 2(5): 882-889.

Masclaux FG, Wyss T, Mateus-Gonzalez ID, et al. (2018) Variation in allele frequencies at the bg112 locus reveals unequal inheritance of nuclei in a dikaryotic isolate of the fungus *Rhizophagus irregularis*. *Mycorrhiza* 28(4): 369-377.

Masclaux FG, Wyss T, Pagni M, et al. (2019) Investigating unexplained genetic variation and its expression in the arbuscular mycorrhizal fungus Rhizophagus irregularis: A comparison of whole genome and RAD sequencing data. *Plos One* 14(12): e0226497.

Mateus ID, Masclaux FG, Aletti C, et al. (2019) Dual RNA-seq reveals large-scale non-conserved genotype x genotype-specific genetic reprograming and molecular crosstalk in the mycorrhizal symbiosis. *ISME Journal* 13(5): 1226-1238.

Mateus ID, Rojas EC, Savary R, et al. (2020) Coexistence of genetically different Rhizophagus irregularis isolates induces genes involved in a putative fungal mating response. *ISME Journal* 14(10): 2381-2394.

Peña R, Robbins C, Cruz Corella J, et al. (2020) Genetically different isolates of the arbuscular mycorrhizal fungus *Rhizophagus irregularis* induce differential responses to stress in cassava. *Frontiers in Plant Science* 11: 1-13.

Quinlan AR (2014) BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* 47(1): 1-34.

Rodriguez-Echeverria S, Teixeira H, Correia M, et al. (2017) Arbuscular mycorrhizal fungi communities from tropical Africa reveal strong ecological structure. *New Phytologist* 213(1): 380-390.

Ropars J, Toro KS, Noel J, et al. (2016) Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nature Microbiology* 1(6): 16033.

Rosikiewicz P, Bonvin J and Sanders IR (2017) Cost-efficient production of in vitro Rhizophagus irregularis. *Mycorrhiza* 27(5): 477-486.

Savary R, Dupuis C, Masclaux FG, et al. (2020) Genetic variation and evolutionary history of a mycorrhizal fungus regulate the currency of exchange in symbiosis with the food security crop cassava. *ISME Journal* 14(6): 1333-1344.

Savary R, Masclaux FG, Wyss T, et al. (2018a) A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus Rhizophagus irregularis. *ISME Journal* 12(1): 17-30.

Savary R, Villard L and Sanders IR (2018b) Within-species phylogenetic relatedness of a common mycorrhizal fungus affects evenness in plant communities through effects on dominant species. *Plos One* 13(11): e0198537.

Schmieder R and Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6): 863-864.

Schmieder R, Lim YW, Rohwer F, et al. (2010) TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* 11: 341.

Scott TW, Kiers ET, Cooper GA, et al. (2019) Evolutionary maintenance of genomic diversity within arbuscular mycorrhizal fungi. *Ecol Evol* 9(5): 2425-2435.

St-Arnaud M, Hamel C, Vimard B, et al. (1996) Enhanced hyphal growth and spore production of the arbuscular mycorrhizal fungus *Glomus intraradicis* in and *in vitro* system in the absence of host roots. *Mycological Research* 100(3): 328-332.

Steidinger BS, Crowther TW, Liang J, et al. (2019) Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* 569(7756): 404-408.

Tatusov RL, Fedorova ND, Jackson JD, et al. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.

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 of the United States of America* 110(50): 20117-20122.

Van der Heijden MGA, Klironomos JN, Ursic M, et al. (1998) Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* 396: 69-72.

Wang L, Wang S and Li W (2012) RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28(16): 2184-2185.

Wang N, Thomson M, Bodles WJ, et al. (2013) Genome sequence of dwarf birch (Betula nana) and cross-species RAD markers. *Mol Ecol* 22(11): 3098-3111.

Wyss T, Masclaux FG, Rosikiewicz P, et al. (2016) Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus *Rhizophagus irregularis*. *ISME Journal* 10(10): 2514-2526.

Yildirir G, Malar CM, Kokkoris V, et al. (2020) Parasexual and Sexual Reproduction in Arbuscular Mycorrhizal Fungi: Room for Both. *Trends Microbiol* 28(7): 517-519.

Zhu YO, Sherlock G and Petrov DA (2016) Whole Genome Analysis of 132 Clinical Saccharomyces cerevisiae Strains Reveals Extensive Ploidy Variation. *G3* 6(8): 2421-2434.

Zou X, Du Y, Wang X, et al. (2020) Genome evolution in Oryza allopolyploids of various ages: Insights into the process of diploidization. *Plant Journal*. Epub ahead of print 2020/11/05. DOI: 10.1111/tpj.15066. https://doi.org/10.1111/tpj.15066.

Supplementary Information

Supplementary figures

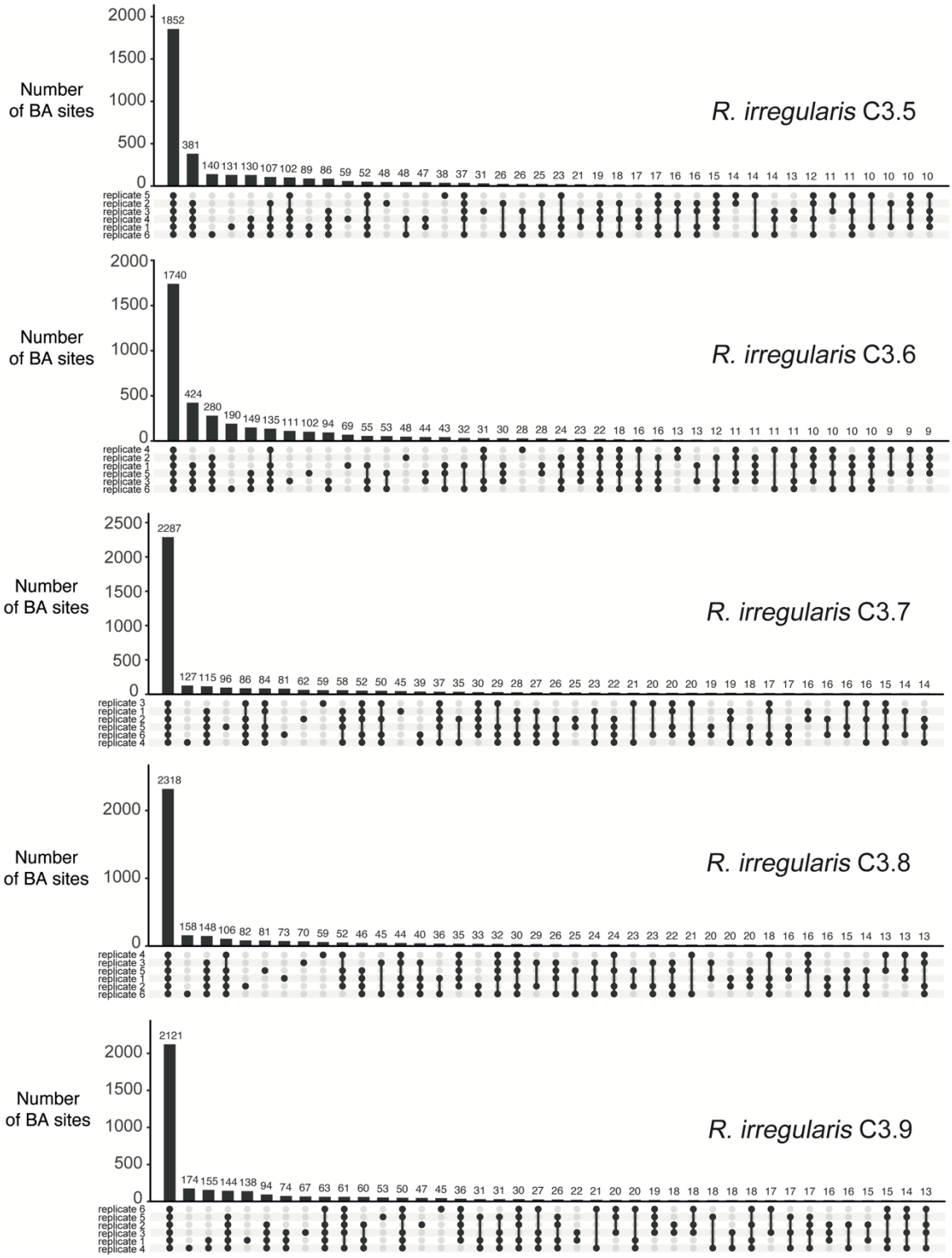


Figure S1: Reproducibility of bi-allelic (BA) sites in the ddRAD-sequencing data among the six replicates in each of the 5 dikaryon Single spores sibling lines (SSSLs) of *R. irregularis* C3. The *y*-axis represents the number of BA sites commonly detected among the replicates being compared. The replicates considered for comparison are labelled at the left and shown along the horizontal axis by darkened circles indicating they were included in the comparison. Isolates are listed from top to bottom and indicated at the far right.

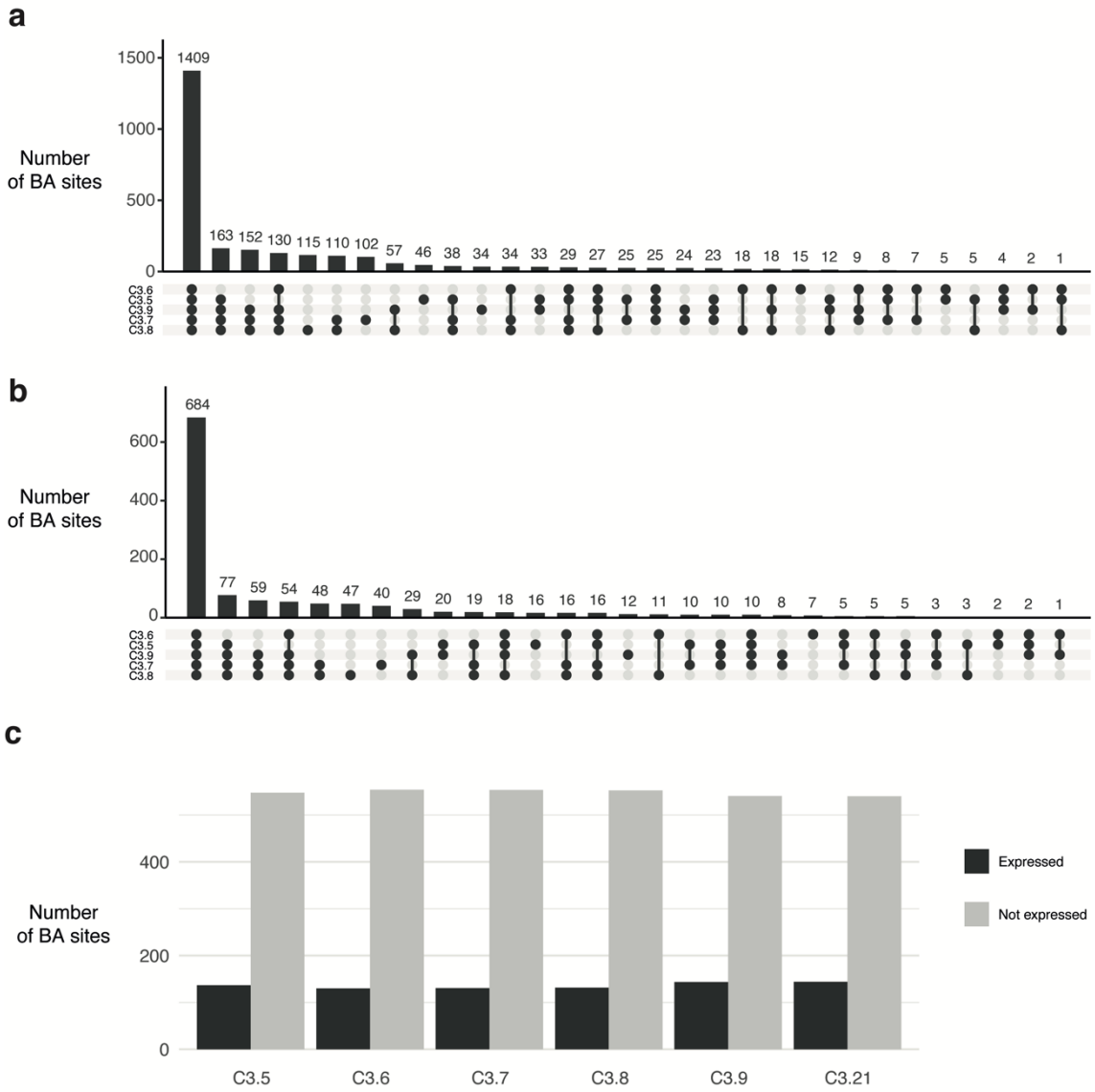


Figure S2: Consistency of bi-allelic (BA) sites in ddRAD-seq data and their transcription among single spore sibling lines (SSSLs) of *R. irregularis* C3. (a) Reproducibility of BA sites in the ddRAD-sequencing data among 5 SSSLs of *R. irregularis* C3. SSSLs being compared are indicated with darkened circles and the number of BA sites is indicated on the *y*-axis. (b) Reproducibility of BA sites within coding sequences in the ddRAD-sequencing data among the 5 SSSLs of *R. irregularis* C3. SSSLs being compared are indicated with darkened circles and the number of BA genes is indicated on the *y*-axis. (c) Evidence of expression of the 684 consistent BA sites from the ddRAD-sequencing data in the RNA-sequencing data.

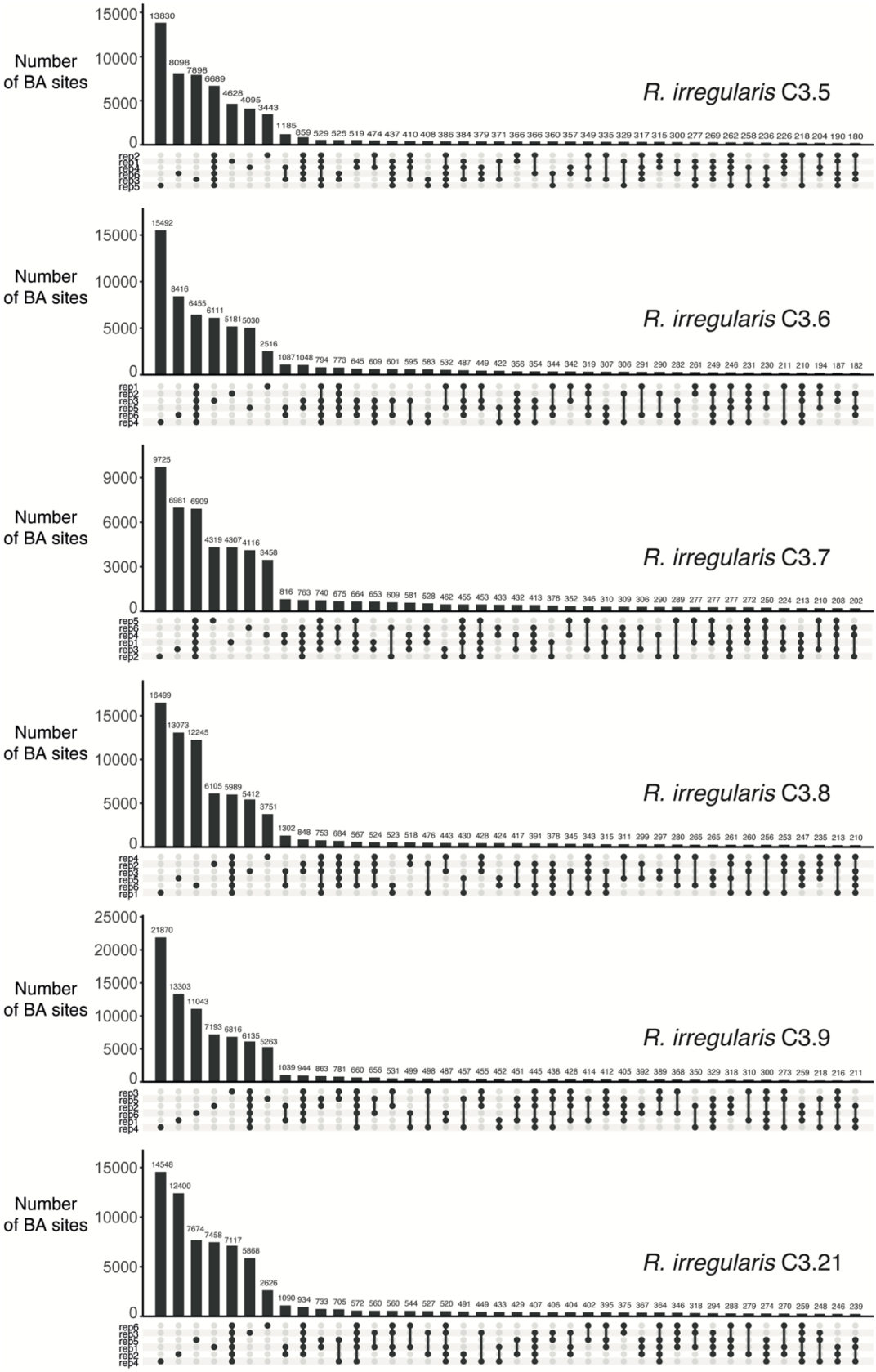


Figure S3: Reproducibility of bi-allelic (BA) sites in the RNA-sequencing data among the six replicates in each of the 6 dikaryon single spore sibling lines (SSSLs) of *R. irregularis* C3. The y-axis represents the number of BA sites commonly detected among replicates being compared. The replicates considered for comparison are labelled at the left and shown along the horizontal axis by darkened circles indicating they were included in the comparison. Isolates are listed from top to bottom and indicated at the far right.

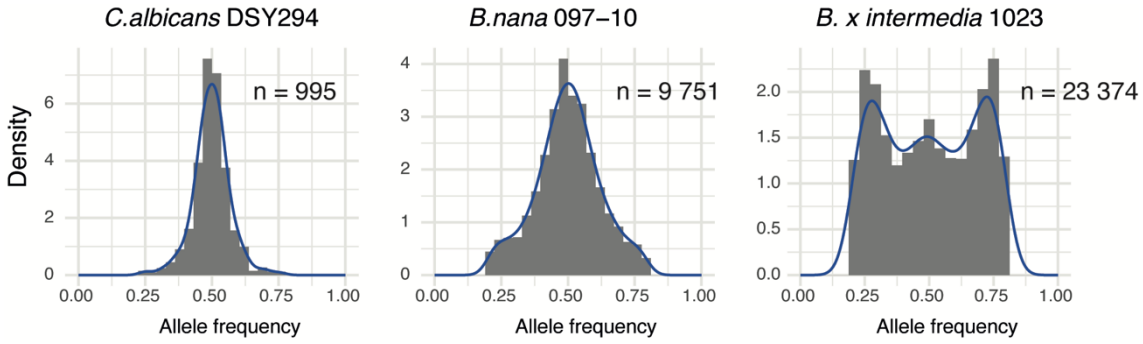


Figure S4: Allele frequency distributions of diploid (*C. albicans* and *B. nana*) and tetraploid (*B. x intermedia*) species derived from ddRAD-seq data. The number of bi-allelic sites analyzed for each species is indicated as (n). Diploid species are expected to exhibit one peak (0.5), and tetraploid species are expected to exhibit three peaks (0.25, 0.5 and 0.75).

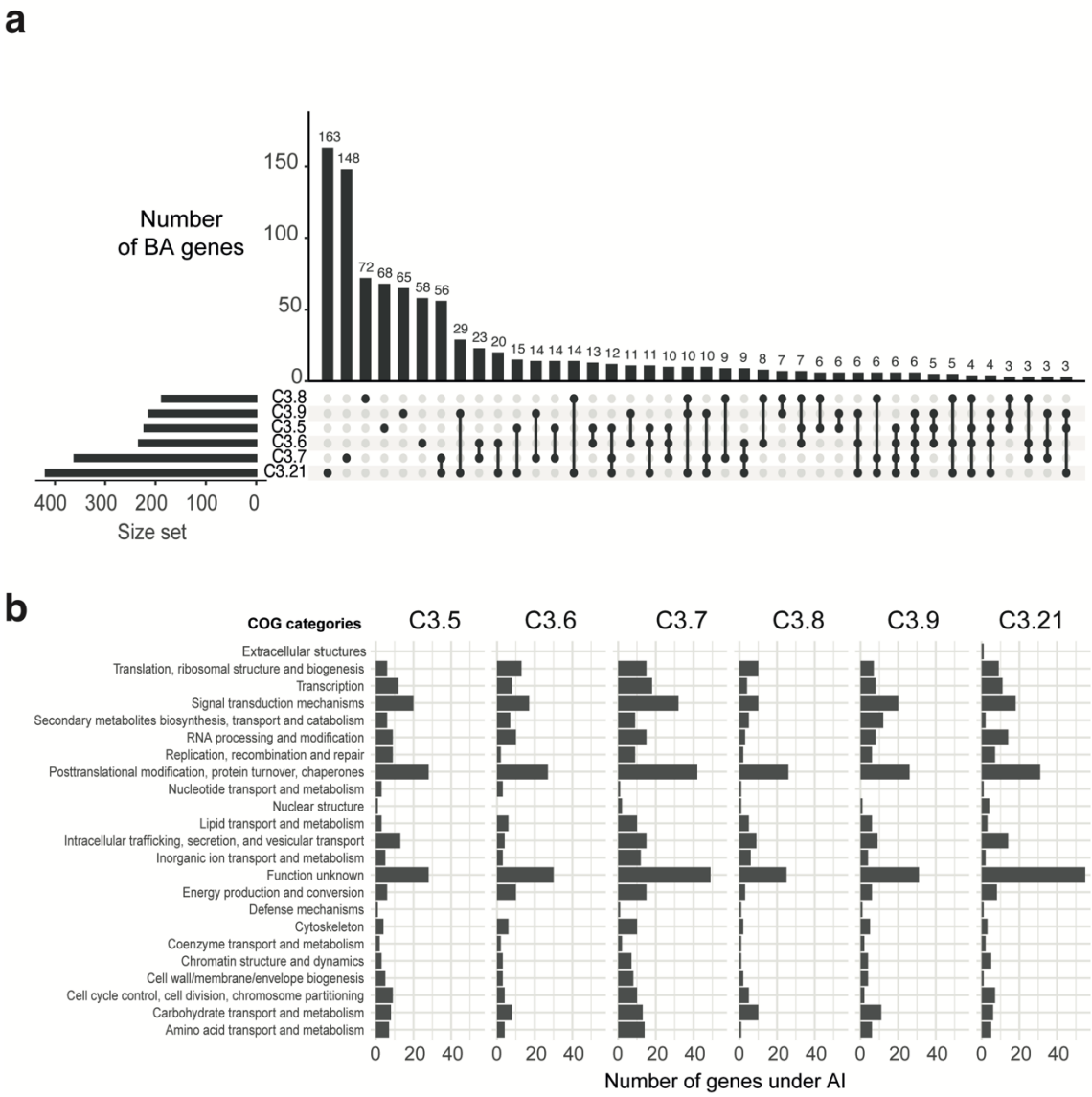


Figure S5: Genes with allele transcription under allelic imbalance. **(a)** Reproducibility of genes under allelic imbalance among the six R. irregularis SSSLs. **(b)** Cluster of orthologous (COG) categories of the genes under allelic imbalance for each *R. irregularis* SSSL.

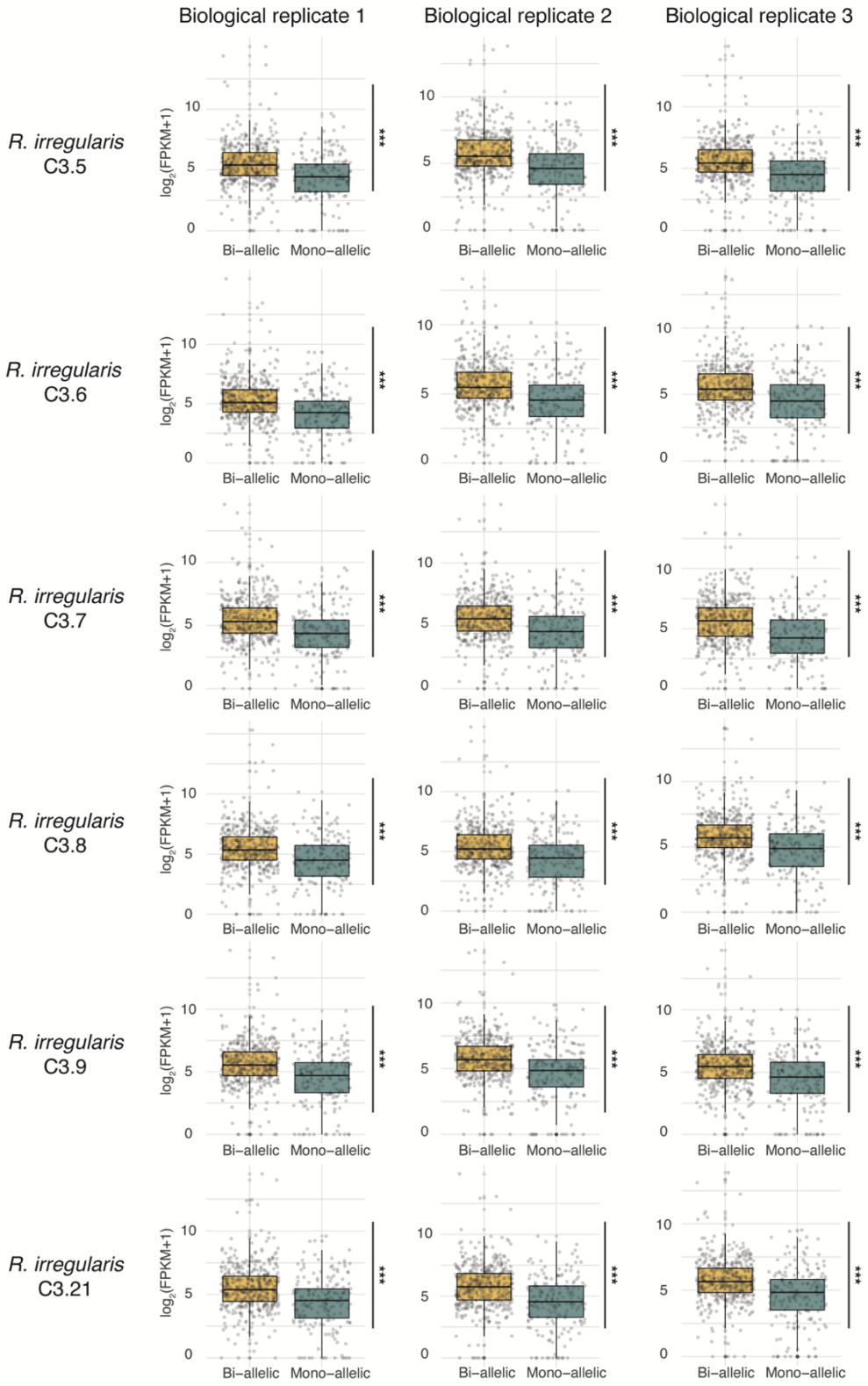


Figure S6: Reproducibility of mono-allelic and bi-allelic expression among three biological replicates of each of the 6 dikaryon single spores sibling lines (SSSLs) of *R. irregularis* C3. Comparison of the expression levels after a $\log_2$(FPKM+1) transformation, between the 187 reproducible genes exhibiting mono-allelic expression and 187 reproducible genes exhibiting bi-allelic expression, for each biological replicate of each *R. irregularis* SSSL. The three asterisks (***) indicate a statistically significant difference between the medians (Mann-Whitney U-test, with a significance threshold of $p < 0.05$). The horizonal line represents the median, the box represents in the interquartile range and the vertical lines represent the maximum and minimum values.

Supplementary tables

All the supplementary tables can be found in a Supplementary MS Excel File in the URL below:

https://nph.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fnph.17530&file=nph17530-sup-0002-TableS1-S11.xlsx

Chapter 4

Clonal dikaryons of the mycorrhizal fungus *Rhizophagus irregularis* generate greater plasticity and variation in molecular responses to changes of the environment than homokaryons

Soon-Jae Lee[1,3], Joaquim Cruz Corella[1,3], Chanz Robbins[1], Gaetan Glauser[2], Ian R. Sanders[1]

[1] Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

[2] Neuchâtel Platform of Analytical Chemistry, Université de Neuchâtel, 2000 Neuchâtel, Switzerland

[3] SJL and JCC contributed equally to this work.

[*]Corresponding author: Ian R. Sanders

This chapter was submitted to *New Phytologist* in February 22nd, 2023.

Abstract

Arbuscular mycorrhizal fungi (AMF) are symbiotic microorganisms forming mutualistic associations with plants. AMF acquire inorganic nutrients that are traded with the plant for carbohydrates and lipids. In the AMF *Rhizophagus irregularis* homokaryons produce genetically identical offspring. Dikaryons produce offspring varying in nucleus genotype proportions. We hypothesized that dikaryon siblings have greater trait plasticity among offspring than homokaryons in response to the environment.

The pre-symbiotic phase is critical in the AMF life cycle. Failure to germinate rapidly and colonize a root is fatal. Strigolactones are signal molecules released by plant roots that influence spore germination. We found strigolactones had negligible effects on homokaryon germination and only on some dikaryon siblings. There was little evidence of conserved AMF metabolomic or transcriptional responses to strigolactone. Proportions of two nucleus genotypes were not affected by strigolactone in dikaryons but strigolactone induced shifts in transcription of two alleles in bi-allelic genes. This occurred in gene sets representing important cellular functions, including genes for alternative splicing.

Our findings demonstrate that trait variation in dikaryons is not dependent on changes in nucleus genotype proportions but can be generated by shifting specific allele expression of bi-allelic genes. This highlights the ability of AMF heterokaryons to produce clonal offspring maintaining high trait variation. This may be key to understanding the enormous ecological success of this widespread symbiont.

Keywords: *Rhizophagus irregularis*, intraspecific variation, strigolactone, alternative splicing, homokaryon, dikaryon, spore germination, metabolite.

Introduction

Arbuscular mycorrhizal fungi (AMF) form widespread symbiotic relationships with most terrestrial plant species (Smith & Read, 2008; Brundrett & Tedersoo, 2018). A mutually beneficial nutrient exchange lies at the foundation of this plant-microbe mutualism. Inorganic nutrients are absorbed by AMF hyphae in the soil. These nutrients are transported to the roots and transferred to the plant. The fungus receives photosynthates and lipids on which it relies for growth (Keymer *et al.*, 2017). This association has enormous impacts on global nutrient cycling, plant growth and plant community structure (van der Heijden *et al.*, 1998; van der Heijden *et al.*, 2003; Koch *et al.*, 2012; Nuccio *et al.*, 2013).

Genetic variation is well known to generate trait variation. Indeed, genetically different AMF vary in quantitative life-history traits and in their effects on plant growth (Angelard *et al.*, 2010; Ceballos *et al.*, 2019; Pena Venegas *et al.*, 2021). Ceballos *et al.* (2019), for instance, demonstrated the relationship between the phylogenetic relatedness, based on patterns of genome-wide single nucleotide polymorphisms (SNPs), of AMF and biomass production of the globally important crop cassava. This indicates that genetic variation among AMF plays a role in their effects on plants.

Genetic variation within the model AMF species, *Rhizophagus irregularis*, is large (Wyss *et al.*, 2016; Savary *et al.*, 2018; Chaturvedi *et al.*, 2021; Robbins *et al.*, 2021). Recent studies showed that while most *R. irregularis* isolates are homokaryons (harbouring one nucleus genotype), some are dikaryons (harbouring two nucleus genotypes) (Ropars *et al.*, 2016; Kokkoris *et al.*, 2021) . This genetic composition, referred to here as karyosis, could likely affect AMF life-history traits, and this is one focus of this study. Single-spore sibling lines (SSSLs), each generated *in vitro* from individual spores of the same parental AMF isolate were shown to be qualitatively indistinguishable in their DNA sequences in terms of presence and absence of SNPs (Angelard *et al.*, 2010; Wyss *et al.*, 2016; Robbins *et al.*, 2021). Thus, a set of SSSLs originating from one parental isolate is considered to represent clonally produced offspring that qualitatively contain the same set of alleles. Clonal siblings of a homokaryon isolate can only inherit nuclei of the same genotype from its parent. Thus, progeny would be identical in their genetic composition. If genetic variation alone determines fungal traits, no significant trait variation would be expected among homokaryon siblings. In contrast, dikaryon SSSLs inherit different proportions of the two parental nucleus genotypes (Masclaux *et al.*, 2018; Robbins *et al.*, 2021). Thus, siblings are qualitatively identical (they each contain the same set of alleles). However, different nucleus genotype proportions mean siblings differ in the proportions of alleles at loci where the two nucleus genotypes carry different alleles (we refer to these hereon as bi-allelic loci). This represents an unusual form of quantitative genetic variation among AMF dikaryon siblings that could result in wider range of trait variation compared to that existing among homokaryon siblings.

We previously showed that generation of unequal proportions of two nucleus genotypes in dikaryon SSSLs is common and influences their transcriptome (Robbins *et al.*, 2021). We found that when

nucleus genotypes were unequal, there was more transcription of an allele located on one nucleus genotype than the other; potentially allowing plasticity in phenotypes among clonal siblings. However, it is unclear how much this imbalance affects AMF traits and whether it can alter AMF responses to the environment. In previous studies, quantitative traits of SSSLs were found to differ (e.g., spore and hyphal densities, plant root colonization), and were shown to differentially influence plant growth in pot experiments and the field (Angelard *et al.*, 2010; Ceballos *et al.*, 2019; Pena Venegas *et al.*, 2021). The magnitude of this variation was large, with up to five-fold differences in rice biomass among plants inoculated with different SSSLs (Angelard *et al.*, 2010), and up to 3-fold in the field with cassava (Ceballos *et al.*, 2019). However, significant differences in cassava growth were also observed among plants inoculated with homokaryon siblings (Ceballos *et al.*, 2019). Homokaryon SSSLs should be genetically identical, suggesting a potential role for epigenetics in trait variation. A previous study found that variation in methylation occurs among the population of nuclei in an AMF, implying generation of epigenetically different homokaryon offspring (Chaturvedi *et al.*, 2021). To our knowledge, the contribution of genetic and epigenetic factors to trait variation among SSSLs is unknown. Consequently, a second goal of this study was to understand how nucleus proportions in dikaryon SSSLs and potentially how epigenetics in homokaryons contributes to trait variation in AMF.

Sanders and Croll (2010) first hypothesized that harbouring more than one nucleus genotype gives rise to trait plasticity in response to the environment. Recently, a study supported this hypothesis, showing that nucleus genotype proportions changed in *R. irregularis* dikaryons in response to the host plant identity (Kokkoris *et al.*, 2021). This is consistent with a previous study by Angelard *et al.* (2014), where host plant identity induced changes in allele frequencies of a nucleus-specific locus (Bg112) among a set of SSSLs. At the time, it was thought that those lines were heterokaryotic with up to four different nucleus genotypes, but subsequent studies in our group showed them to have been dikaryons. While those two studies appear to show that nucleus genotype proportions change in response to the biotic environment, it is unclear if this is because of an active regulation of the nucleus ratios; or whether certain spores with a pre-set stable configuration of nuclei are more prone to successfully thrive in a specific environment, resulting in a progressive alteration of the nucleus ratios to that environment. However, we previously found different nucleus genotype proportions in three of six dikaryon SSSLs cultured in the same environment for over ten years (Masclaux *et al.*, 2018; Robbins *et al.*, 2021). If dikaryon *R. irregularis* isolates altered nucleus genotype ratios in response to the environment, we would expect those ratios to converge to similar values. Therefore, it is paramount to investigate more precisely how traits respond in homokaryon and dikaryon SSSLs to a changing environment and what drives trait response. Differential trait response among genetically identical homokaryons would point to an epigenetic controlled acclimation to the environment. Differential trait responses among dikaryon SSSLs could be epigenetic and/or genetic; where the genetic contribution to trait variation could potentially occur because of a change in nucleus genotype proportions in response to the environment.

Establishment of the AM symbiosis depends on germination of spores. Subsequent hyphal branching increases the likelihood of symbiosis formation. These processes are induced in response to plant-secreted strigolactones (Besserer *et al.*, 2006; Lanfranco *et al.*, 2018). Strigolactones enhance fungal mitochondrial activity, ATP production and gene expression (Besserer *et al.*, 2006; Tsuzuki *et al.*, 2016; Lanfranco *et al.*, 2018). With limited resources, spores must find a host and form the symbiosis within approximately 3 weeks. Therefore, response to strigolactone is a directly fitness-related trait and any variability in response among SSSLs is critical for fitness.

Here, we quantified variation in this trait among homokaryon and dikaryon SSSLs in response to strigolactone exposure. Firstly, we hypothesized that trait variation arises among homokaryon SSSLs in response to strigolactone exposure because of epigenetic variation; and that additional trait variation will be observed among dikaryon SSSLs due to differences in nucleus genotype proportions. This trait variation will be reflected in spore germination rates, gene transcription and secreted fungal exudates. Secondly, we hypothesized that nucleus genotype ratios in dikaryons will shift in response to strigolactone exposure.

Two complementary experiments were performed to test these hypotheses (Fig. 1). Four homokaryon and four dikaryon SSSLs were treated with strigolactone or not treated (Control). In experiment 1, we measured spore germination and assessed fungal exudate profiles. Additionally, we tested whether nucleus genotype ratios shifted in response to strigolactone. In experiment 2, we measured spore germination at three different times and assessed fungal exudate profiles. We also conducted transcriptome profiling to characterize whether differential gene transcription occurred among SSSLs and in response to strigolactone. To discriminate the confounding contribution of epigenetic and quantitative genetic variation to trait variation among siblings, we included homokaryon SSSLs. Because nuclei are genetically identical in homokaryon SSSLs, trait variation among siblings would not be attributable to genetic differences, and would require investigation for a likely epigenetic source.

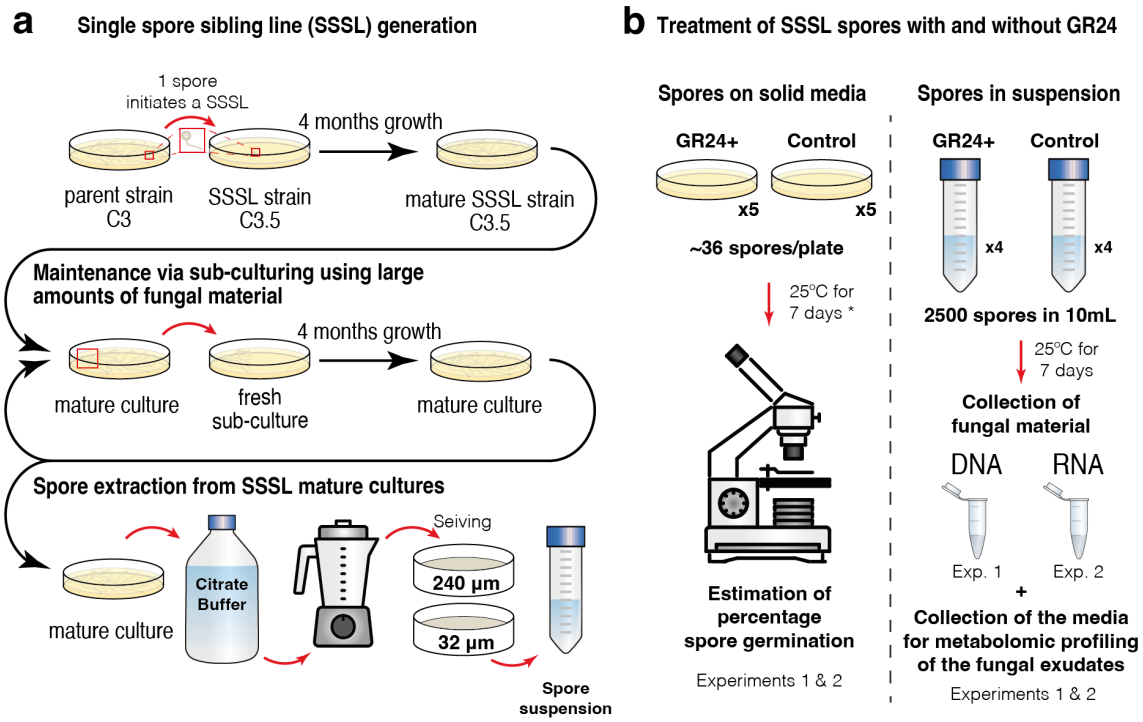


Fig. 1. Production of single spore sibling lines (SSSLs) and the experimental setup. a) Procedure to produce and maintain SSSLs *in vitro* and the procedure to isolate a suspension of spores from mature cultures. b) The experimental setup. On the left side, ~36 spores were placed on each plate (with 5 replicate plates per treatment) and their germination was observed under the microscope. On the right side, suspensions of spores incubated with or without GR24 at 25ºC for 7 days (4 replicates). * In experiment 2, spore germination observations were also made at 14 day and 21 days. DNA/RNA from the fungal material was then extracted, and liquid samples from the media were used for the profiling of fungal exudates. GR24-treated samples are labelled (GR24+), non-treated samples are labelled (Control). "Blender" and "Microscope" icons by Dooder and Lastpark, from thenounproject.com.

Materials and Methods

We conducted an experiment twice with the same design to check for reproducibility of trait responses to strigolactone and to test the hypotheses regarding a shift in nucleus genotype proportions in response to strigolactone (experiment 1); transcription profile responses to strigolactone (experiment 2). Limited fungal material meant this could not be performed in one experiment.

Fungal material and culturing

We used 4 *R. irregularis* homokaryons (C2, C2.5, C2.6 and C2.8), and 4 *R. irregularis* dikaryons (C3.5, C3.6, C3.9, C3.21). C2.5, C2.6 and C2.8 were produced by the parental isolate C2. C3.6, C3.9 and C3.21 were produced by C3.5, that is itself a SSSL of isolate C3 (Robbins *et al.*, 2021). Dikaryon SSSLs were chosen based on their nucleus genotype proportions (Masclaux *et al.*, 2018; Robbins *et al.*, 2021). C3.5 and C3.6 had almost equal proportions of both nucleus genotypes, and C3.9 and C3.21 displayed unequal proportions biased towards the same nucleus genotype. None of these SSSLs differed qualitatively from the parent meaning that they share the same alleles (Robbins *et al.*, 2021). All the SSSLs were produced *in vitro* from individual spores of the respective parental isolate, as described by Angelard *et al.* (2010), and maintained at 25 °C in axenic conditions with Ri T-DNA-transformed carrot roots with sub-culturing every 4 months (Fig. 1a).

After 4 months, fungal spores from mature cultures were collected by dissolving the medium from petri dishes in 2 L of stirred citrate buffer (0.0062 M citric acid, 0.0028 M sodium citrate) for 20 minutes, blended for 30 seconds, then filtered (Fig. 1a). Two sieves (pore size 200 μm and 32 μm) were used to filter out root tissue and fungal hyphae. Spores were rinsed with sterile distilled H_2O, collected and kept at 4 °C for 7 days.

Experimental set-up

GR24 is a synthetic analogue of a strigolactone that is known to induce a similar response in the fungus to that of the presence of a root. It is widely used in studies on pre-symbiotic AMF growth (Besserer *et al.*, 2006; Tsuzuki *et al.*, 2016). Two strategies were used in each experiment to test strigolactone (GR24) treatment effects on the fungus (Fig. 1b).

Firstly, we prepared 5 Petri plates (representing 5 replicates) per SSSL with water-agar media containing 0.01 μM GR24, and 5 plates without GR24. Thirty-six spores were then placed on each plate and incubated at 25 °C for 7 days. In experiment 1, after 7 days, we counted the total number of germinated spores on each plate and calculated the percentage of germinated spores. In experiment 2, we followed the same procedure but calculated the percentage of germinated spores at 7, 14 and 21 days.

Secondly, we prepared spore suspensions in tubes, each containing ~2500 spores in 10 mL of sterile distilled H_2O. Each tube represented one replicate. In both experiments, 4 replicates of each

homokaryon SSSL were treated with 10 µL of the strigolactone GR24 solution (0.01 µM), and 4 replicates of each homokaryon SSSL with an equivalent volume of H_2O. Because more dikaryon material was available, we prepared 6 replicates and 8 replicates of each SSSL per treatment in experiments 1 and 2, respectively. All samples were kept in the dark at 25 °C for 7 days. We then filtered the samples (40 µm pore size; Corning® Cell Strainer) to retain the fungal material for DNA or RNA extraction. The liquid was retained for metabolic profiling of fungal exudates. Fungal material was flash frozen and stored at -80 °C. This was done in the same way in both experiments.

Metabolic profiling

Metabolome analysis of fungal exudates was carried out using an Acquity UPLC coupled to a Synapt G2 Quadrupole - Time of Flight (QTOF) mass spectrometer (UHPLC-QToF-MS; Waters). An Acquity UPLC HSS T3 column (100x2.1mm, 1.8 µm; Waters) was used with a flow rate of 500 µl/min and maintained at 40°C. A gradient with 0.05% formic acid in water as mobile phase A and 0.05% formic acid in acetonitrile as mobile phase B was applied: 0-100 % B in 10 min, holding at 100% B for 2.0 min, re-equilibration at 0% B for 3.0 min. The injection volume was 2.5 µl. The QTOF was operated in electrospray positive mode using the MSE acquisition mode that alternates between two acquisition functions; one at low and another at high fragmentation energies. Mass spectrometric parameters were: mass range 50-1200 Da, scan time 0.2 s, source temperature 120°C, capillary voltage 2.8 kV, cone voltage 25V, de-solvation gas flow and temperature 900 L/h and 400°C, respectively, cone gas flow 20 L/h, collision energy 4 eV (low energy acquisition function) or 8-55 eV (high energy acquisition function). A 500 ng/ml solution of the synthetic peptide leucine-enkephaline in water:acetonitrile:formic acid (50:50:0.1) was infused constantly into the mass spectrometer as internal reference to ensure accurate mass measurements (<2ppm). Data was recorded by Masslynx v.4.1. Marker detection was performed using Markerlynx XS (Waters) with the parameters: initial and final retention time 1.0 and 10.0 min, mass range 85-1200 Da, mass window 0.02 Da, retention time window 0.08 min, intensity threshold 500 counts, automatic peak width and peak-to-peak baseline noise calculation, deisotoping applied. Data was mean-centred and Pareto scaled before applying multivariate analysis.

Nucleus genotype proportions

Fungal material from experiment 1 was homogenized (CryoMill; 2 x 30 s, 25 Hz, resting 30 s, 5 Hz) and used to extract DNA (Qiagen Plant DNA Kit). The DNA was quantified (Promega QuantusTM Fluorometer and DNA QuantiFluor® dye) and stored at -20 °C. Two alleles of a putative MAT locus (MAT-1 and MAT-2) are nucleus genotype-specific in *R. irregularis* dikaryons. We used quantitative polymerase chain reaction (qPCR) with specific primers targeting the two alleles, MAT-1 and MAT-2, in the dikaryon SSSLs. This allowed us to estimate nuclear genotype proportions of each SSSL in each treatment. The CYBR green qPCR was performed under the conditions: 10 mins denaturation at 95°C, followed by 38 cycles of 95°C for 30 s, 56°C for 30 s, and 70°C for 30 s, with final elongation at 72°C for 10 min. Melting curve analysis was performed following the last cycle of qPCR. Primer sequences and their melting temperatures (Tm) are shown in Table S1. Primer efficiencies and specificities are shown in Fig. S1. Non-template controls with each primer pair were used to check for non-target amplification. All qPCRs performed with 6 biological replicates with 3 technical replicates. Differences in the quantification cycle (Cq) values of the two MAT-like alleles (ΔCq(MAT-1, MAT-2)) were used to determine the nucleus genotype proportions.

RNA sequencing

RNA was extracted in experiment 2 (Maxwell RSC Plant RNA kit; Promega) from homogenized flash-frozen spores (CryoMill; Retsch GmbH, Haan, Germany) (2× 30 s, 25 Hz, resting 30 s, 5 Hz). RNA quantity and quality were determined (Nanodrop photometer and Agilent 5200 Fragment Analyzer). Three replicates were prepared of each SSSL in each of the treatments.

Forty-eight libraries (8 SSSLs x 2 treatments x 3 replicates) were prepared using 100 ng of total RNA. Library construction followed the TruSeq protocol: TruSeq® Stranded mRNA Library Prep (Illumina, San Diego, CA, USA). To generate fragments for paired-end sequencing, fragmentation time was reduced to 1 min (from the standard 8min). Adapter-ligated cDNA fragments were enriched by 17 cycles of PCR using the Illumina TruSeq PCR Primer Cocktail. Final libraries were purified twice with SPRIselect beads (Beckman Coulter) using a ratio of 0.7 and 1, respectively.

Sequencing was performed on a NovaSeq 6000 System (Illumina, San Diego, CA, USA) at the Lausanne Genomic Technologies Facility. Demultiplexed data files were deposited with European Nucleotide Archive (accession PRJEB52540).

RNA-sequence data

Quality control of the sequences was carried out using FASTQC (v.0.11.4) (Andrews, 2010). Sequences matching the Illumina TruSeq adapters were removed, and low-quality bases ($Q < 24$) found towards the ends of reads were trimmed using TRIMMOMATIC (v.0.30) (Bolger *et al.*, 2014). Trimmed reads were aligned to their reference genome using STAR (v.2.7.8a) (Dobin *et al.*, 2013). Sequences from homokaryons were mapped to the parental C2 genome assembly (NCBI accession: GCA_903189765.1). Dikaryons were mapped to the parental C3 genome assembly (NCBI Accession:

GCA_903189775.1) that is partially phased (Chaturvedi *et al.*, 2021). Only reads mapping to a single location were retained for downstream analyses. Mapping summary statistics are presented in Table S2.

Allele frequency analysis

Analysis of allele frequencies of transcripts at bi-allelic sites was carried out as described by (Robbins *et al.*, 2021). Bi -allelic sites were called using FREEBAYES (v.1.3.5) (Garrison & Marth, 2012) with the following parameters: -p 2 -J -K -F 0.1 -0 -u. Allele frequencies were estimated at each bi-allelic site by pooling the number of reads that carried each allele in each replicate using a custom python script (https://github.com/jquimcrz/afreq_NGS). To detect GR24 induced shifts in allele frequency at each bi-allelic gene, we performed pairwise Welch's t-tests, and all probabilities were adjusted for multiple comparisons using a Benjamini-Hochberg (BH) correction. Genes with adjusted $p < 0.05$ were considered to have a significant shift and were used for gene set enrichment analyses with String database v11.5 (von Mering *et al.*, 2003). Pathways with FDR < 0.05 were considered significantly enriched.

Differential gene expression

The number of reads aligning to each gene was counted using 'htseq-count' in HTSeq (v.0.11.2) (Anders *et al.*, 2015) with the parameters: --stranded=no -r pos -f bam. Differential gene expression analysis was conducted using 'DESeq2' (v.1.34.0) (Love *et al.*, 2014). Differential expression was tested between the GR24+-treated and controls, in the two sets SSSLs separately. Genes with a minimum sum of 10 reads across all samples were considered in the analyses. Genes with an absolute $\log_2$(Fold Change) > 1 and an FDR < 0.05 were considered differentially expressed.

Transcription of known GR24-upregulated genes

By finding the reciprocal best hits using 'blastp' (McGinnis & Madden, 2004), we identified the corresponding sequences of 19 genes in assemblies isolates C2 and C3 (Chaturvedi *et al.*, 2021) that were previously shown to be upregulated in *R. irregularis* DAOM197198 in response to GR24 (Tsuzuki *et al.*, 2016) (Tables S3 and S4). We checked transcription levels of these genes in the RNA-seq data to test whether upregulation of those genes consistently occurred in the presence of GR24 in all *R. irregularis* SSSLs (Tables S5 and S6).

Statistical analyses

All the statistical analyses have been carried out using R (v. 4.0.3) (R Core Team, 2020). Differences in mean percentages of germination among SSSLs were tested using Welch's unequal variances *t*-tests, and all *p*-values were adjusted for multiple comparisons using Benjamini-Hochberg (BH) correction. Differences in the median $\Delta Cq_{(MAT\text{-}1,\ MAT\text{-}2)}$ values between the GR24 treatments and untreated controls were tested using the non-parametric Mann-Whitney *U* test for each dikaryon SSSL

independently, and *p*-values were adjusted for multiple testing using BH correction. To test differences in $\Delta Cq_{(MAT\text{-}1,\ MAT\text{-}2)}$ among the four dikaryon SSSLs we performed a Kruskal-Wallis rank sum test, followed by a post-hoc analysis using the Mann-Whitney *U* test to perform the pairwise comparisons. All *p*-values were adjusted for multiple testing using a BH correction. Welch's unequal variances *t*-tests and Mann-Whitney *U* tests were performed with the base R functions 't.test' and 'wilcox.test' respectively. Kruskal-Wallis test was performed using the base R function 'kruskal.test'. Hierarchical clustering of the LC-MS samples was performed using the R package "pheatmap" by computing Euclidean distances and clustering using the Ward's method ("ward.D2"), after scaling the data. Non-metric Multidimensional Scaling (NDMS) of the LC-MS data was performed independently for the homokaryon and dikaryon SSSL families using the 'metaMDS' function of the "vegan" (v. 2.4-2) R package (Oksanen et al., 2020), using the Euclidean distance metric, for a number of dimensions k=2 and 50 random restarts. Permutational multivariate analysis of variance of LC-MS data was performed using the same "vegan" R package. Firstly, Euclidean distance matrices were computed using the 'vegdist' function; then, the analysis of the variance was performed with the 'adonis' function (999 permutations) to test the effect of the GR24 treatment and of the SSSL identity. Post-hoc statistical tests were carried out with the 'pairwise.adonis' function of the "Pairwise.Adonis" (Martinez Arbizu, 2020) R package (*p*-values were BH adjusted for multiple testing). All the figures were produced using the package "ggplot2" (Wickham, 2016), except for the Venn diagrams that were generated using an online tool available at: http://bioinformatics.psb.ugent.be/webtools/Venn/.

Results

Dikaryon SSSLs germinated more rapidly than homokaryon SSSLs

In both experiments, we observed that dikaryons displayed higher percentage of spore germination after 7 days compared to that of homokaryons ($t = -34.889$, $df = 70.702$, $p < 2.2e^{-16}$ and $t = -9.8308$, $df = 48.136$, $p = 4.282e^{-13}$, in experiments 1 and 2, respectively; Fig. S2). These differences were observed between the two SSSL families irrespectively of the GR24 treatment (Fig. S2). While all dikaryon SSSLs exhibited mean percentages of germination of > 90% in both experiments, homokaryon SSSLs ranged from 60% to 70%.

In experiment 2, spore germination was also counted over time, revealing that homokaryons took longer to germinate (Fig. S3). By 21 days, all SSSLs reached a mean percentage of germination above 90%, at which time there were no significant differences between dikaryons and homokaryons.

Spore germination in response to GR24

No significant differences in spore germination were observed in the homokaryon SSSLs in response to GR24 in either experiment (Fig 2). In both experiments, the dikaryon SSSLs C3.9 and C3.21 exhibited significantly higher spore germination percentages in GR24 treatments than in untreated controls, with a difference of approximately 10% (Fig. 2). GR24 treatment did not affect spore germination in SSSLs C3.5 and C3.6 in either experiment.

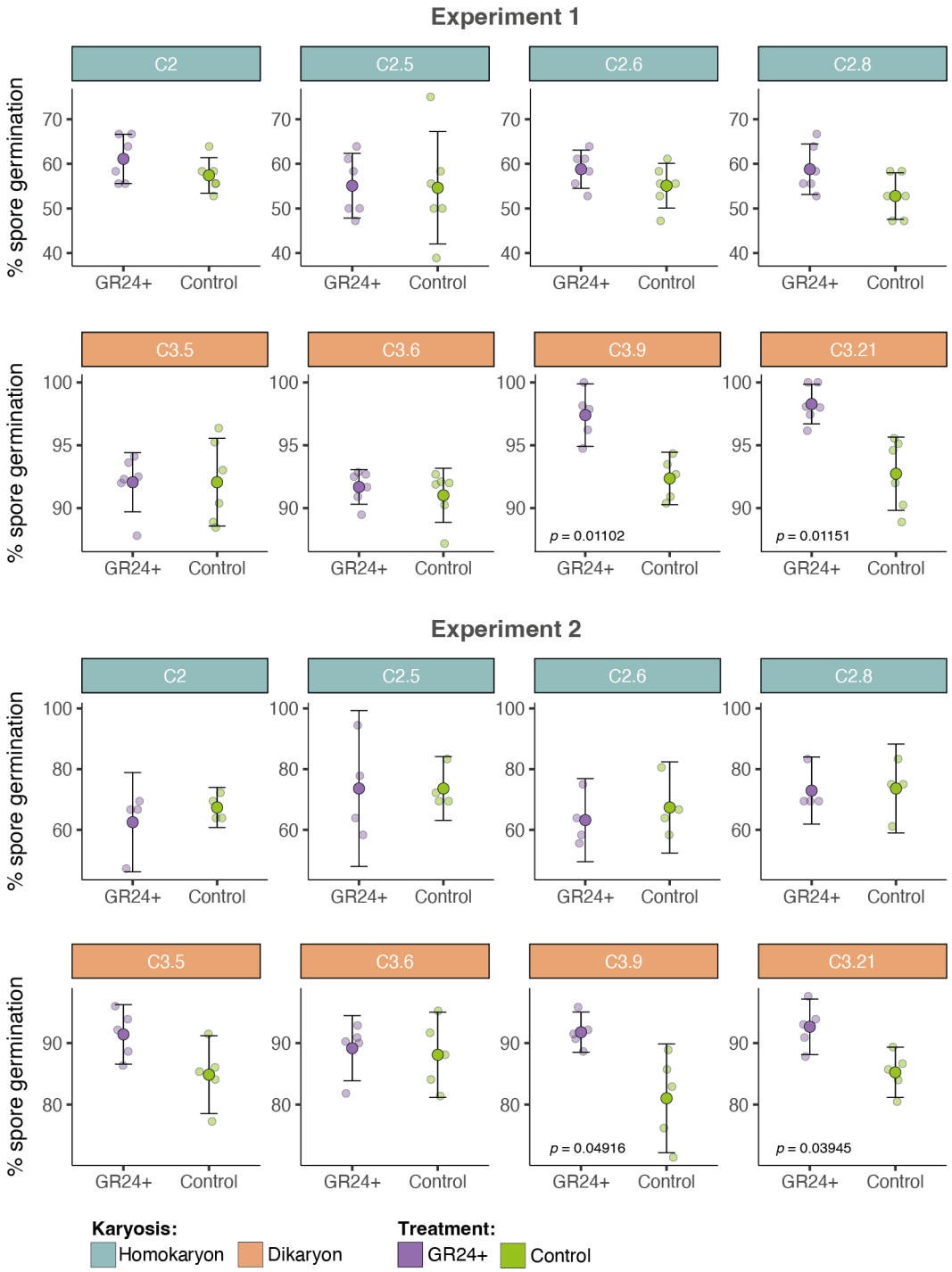


Fig. 2. Spore germination among SSSLs in two independent *in vitro* experiments. The percentage of germinated spores of each R. irregularis SSSL is shown for GR24 treated (GR24+) and untreated (Control) samples 7 days after being placed on the soild medium. Individual replicates are represented by semi-transparent data points. Soild dots and error bars represent the mean percentages of germination, and the bars represent the 95% confidence interval. Probability (*P*) of the test for a GR24 treatment effect on spore gemination, following Benjamini-Hochberg correction for multiple testing, is shown where a significant effect was observed.

Fungal exudate profiles differed between homokaryons and dikaryons and were affected by strigolactone

Hierarchical clustering based on the exudate metabolome profiles of each SSSL revealed a clear separation between the homokaryons and dikaryons (Fig. 3). This was consistent in the two experiments. Inner clustering was also observed due to GR24 treatment and the SSSL identity. This suggested additional effects of these factors on the metabolome. Inner clusters indicate the presence of SSSL-specific and treatment-specific metabolites in both SSSLs families.

We identified the presence of metabolites specific to each treatment and to each SSSL (Fig. S4). In experiment 1, differences occurred among homokaryon SSSLs in the number of metabolites consistently found in all replicates of each SSSL and between treatments (Fig. S4a). However, variation was more pronounced among the dikaryon SSSLs and in response to GR24 than among homokaryon SSSLs. C3.5 and C3.6 displayed a larger number of GR24-responsive metabolites (> 150) compared to controls and to C3.9 and C3.21. Only a few GR24-specific metabolites were commonly found among all SSSLs of the same family (10 in homokaryons; 24 in dikaryons) (Fig. S4b, Fig. S5). Among these, only 7 GR24-responsive metabolites were found in all homokaryon and dikaryon SSSLs in experiment 1. No control-specific metabolites were commonly found in the two families. In experiment 2, the number of SSSL-specific metabolites between treatments also varied in homokaryons and dikaryons (Fig. S4a). Among these, only one GR24-specific metabolite was commonly detected in all homokaryon SSSLs. In dikaryons, 13 GR24-specific metabolites were detected (Fig. S4b). No control-specific metabolites were commonly detected in both SSSL families (Fig. S4b). This variability in the presence/absence of treatment specific metabolites in experiment 2 was reflected in the lower number treatment-specific metabolites commonly found in all siblings, and metabolites shared between the two SSSL families compared to those found in experiment 1 (Figs. S4b). Most notably, the number of SSSL-specific metabolites detected in both experiments, and in both SSSL families, was higher than the number of treatment-specific metabolites among SSSLs (Fig. S5).

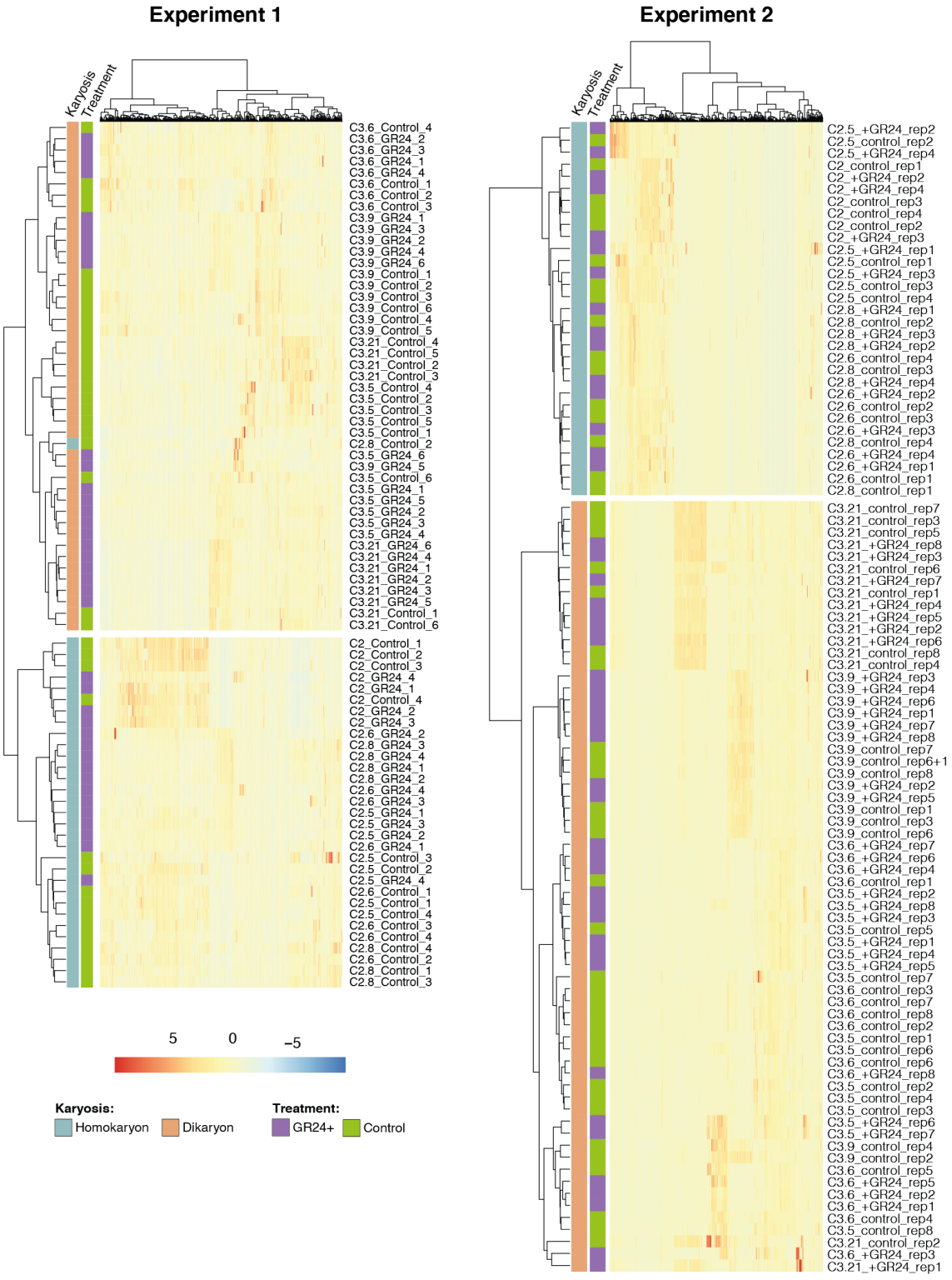


Fig. 3. Hierarchical clustering of the metabolome profiles obtained by LC-MS form the SSSLs in the two experiments. The colour gradient bar (-5 to 5) represents the scaled abundance of the different metabolites.

The similarities among siblings of each SSSL family in response to GR24 were summarized using NDMS (Fig. 4). Differences between the two experiments were reflected in the distribution of the samples, with a more obvious effect of GR24 in experiment 1, where treated and control samples of each SSSL separated along the NMDS2 axis. The GR24 effect in experiment 1 was significant in both

SSSL families (homokaryons $p < 0.001$; dikaryons $p < 0.001$) (Tables S7a and S9a). In experiment 2, GR24-treated and untreated samples formed more homogenous clusters in the NMDS plots. The treatment effect was significant among dikaryon SSSLs ($p = 0.019$), but not among homokaryons (Tables S8a and S10a).

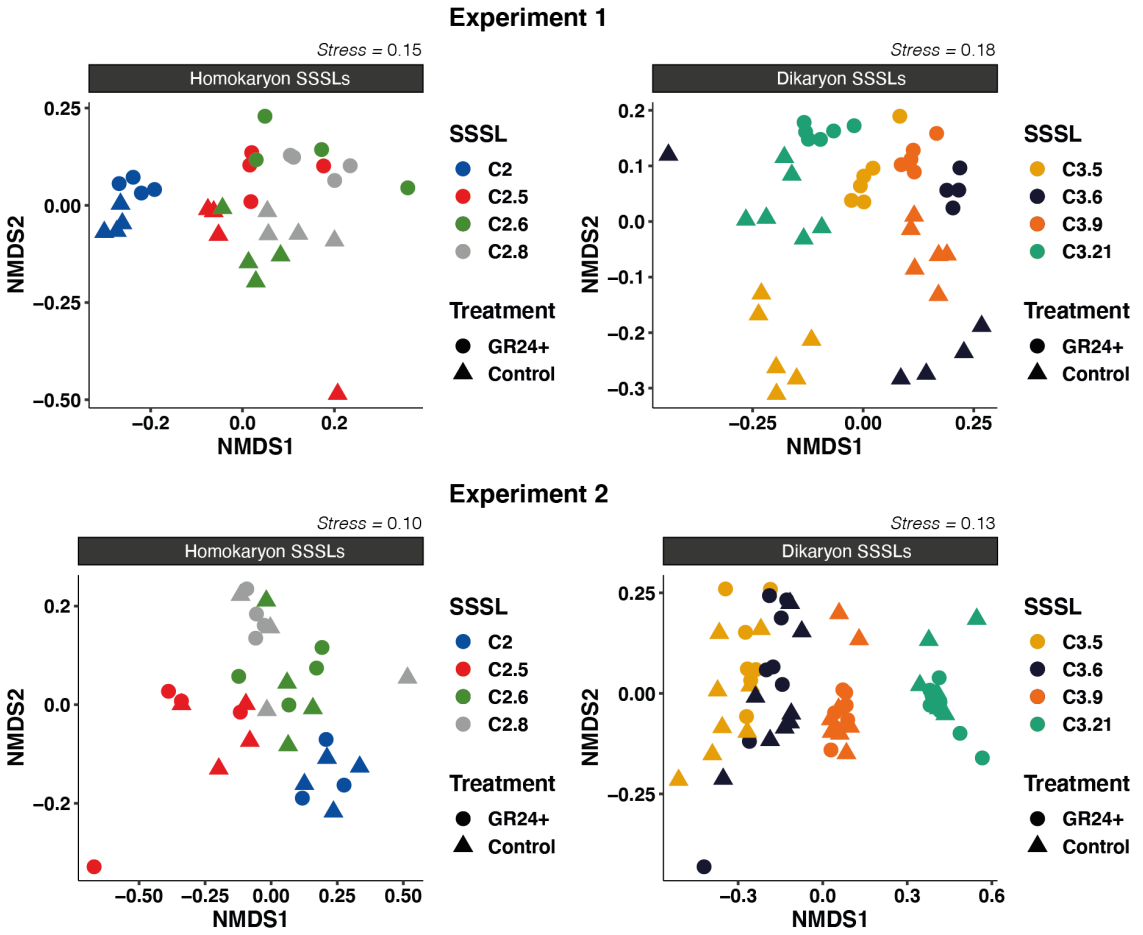


Fig. 4. Non-metric multi-dimensional scaling (NMDS) of the fungal metabolome data generated by LC-MS of each SSSL in each treatment. The stress value represents the goodness of NMDS. The lower the stress value, the better fit of original distances/dissimilarities and projected distances in ordination diagram is reached.

Additionally, we tested the null hypothesis that the relative abundance of metabolites does not differ among SSSLs of the same parent. Homokaryon siblings did not display different metabolome profiles in experiment 1, but displayed different metabolome profiles in experiment 2 ($p < 0.001$) (Tables S7a and S7b-c; and S9a). In experiment 2, differences were observed among all SSSLs regardless of the treatment (Table S9b). The relationship among dikaryon SSSL metabolomes was more consistent between the two experiments. In experiment 1, metabolome profiles of C3.5 differed from those of C3.9 and C3.21 when not exposed to GR24 ($p < 0.05$); and metabolome profiles of C3.5 and C3.6 significantly differed from those of C3.21 when treated with GR24 ($p < 0.05$) (Tables S8b-c). In experiment 2, metabolome profiles of C3.21 differed from those of C3.5 and C3.6 when not treated with GR24 ($p < 0.05$); and metabolome profiles of C3.21 and C3.9 differed from those of C3.5 and C3.6 when treated with GR24 ($p < 0.05$) (Tables S10b-c).

Nucleus genotype proportions were not affected by GR24

The relative proportion of two nucleus-specific alleles (MAT-1 and MAT-2) at the putative MAT-locus indicated that variation in nucleus genotype proportions existed among dikaryon SSSLs (Fig. 5a). C3.5 and C3.6 exhibited a higher proportion of MAT-1. In contrast, MAT-2 was dominant in C3.9 and C3.21. No dikaryon SSSLs displayed significant differences in MAT-1:MAT-2 ratios between GR24 and control samples, meaning that GR24 did not alter nucleus genotype proportions.

Allele frequencies at bi-allelic loci

SSSLs C3.9 and C3.21 displayed a bi-modal allele frequency distribution at bi-allelic loci, with peaks close to 0.3 and 0.7. The bi-modal allele frequencies in C3.9 and C3.21 were observed in the two SSSLs that displayed a higher abundance of the MAT-2 nucleus genotype.

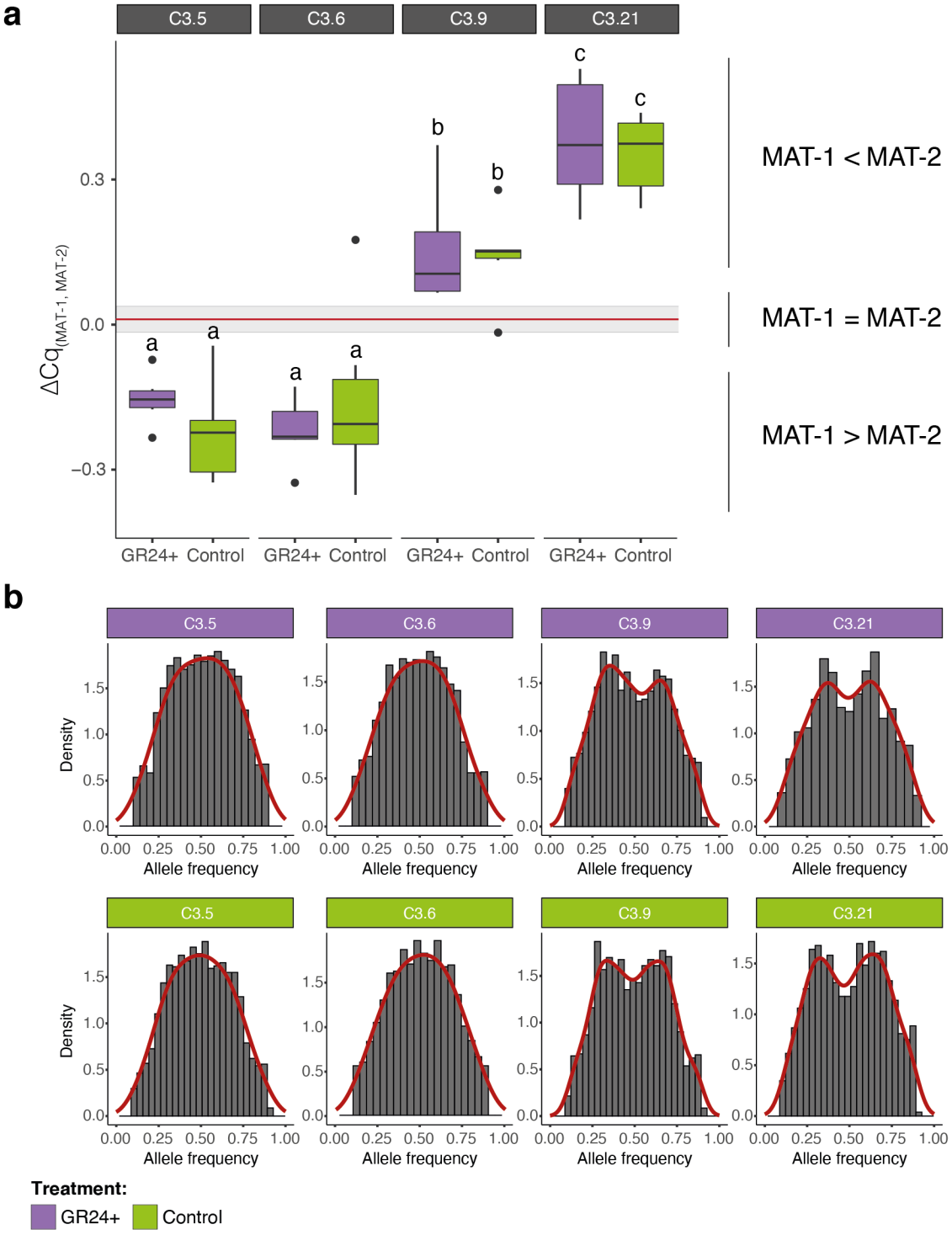


Fig. 5. Nucleus genotype proportions among dikaryon SSSLs. a) Estimation of the relative abundance of two nucleus genotypes among dikaryon SSSLs and across experimental conditions assessed by qPCR assays of two nucleus genotype-specific alleles (MAT-1 and MAT-2) of the putative MAT-locus. ΔCq(MAT-1, MAT-2) values were obtained by subtracting the Cq of MAT-2 from the Cq of MAT-1. Negative ΔCq(MAT-1, MAT-2) values indicate the dominance of the MAT-1 nucleus genotype-specific allele, whereas positive values indicate the dominance of MAT-2 nucleus genotype-specific allele. The horizontal red line represents the mean ΔCq(MAT-1, MAT-2) of triplicates of an artificial 1:1 mixture of MAT-1:MAT-2, and the grey area is its standard deviation. Different letters above boxplots represent statistically significant differences ($p \leq 0.05$ after 'Benjamini-Hochberg correction for multiple testing). b) Distribution of allele frequencies at bi-allelic sites in the transcriptome data of dikaryon SSSLs.

Transcriptome profile differences among SSSLs

Transcriptome profile differences were found among SSSLs of both families, as shown by the clear clusters formed by the SSSLs, after performing a principal component analysis on the transcript data (Fig. 6a). Differences between the treatments were smaller than that observed among the SSSLs. We further identified the existence of several SSSL-specific differentially expressed genes (Fig. 6b, Tables S11b-e and S12b-e). Most notably, many more SSSL-specific genes were differentially expressed as a result of exposure to GR24 than those common to all SSSLs of a family. This indicates that there is not a strongly conserved transcriptional response to strigolactones, even within an AMF species. Notably, there was a larger number of SSSL-specific, GR24 responsive genes observed in the dikaryons (75 SSSL-specific genes) than among homokaryons (21 SSSL-specific genes) (Fig. 6b).

Transcriptome differences among SSSLs

Transcriptome profile differences were found among SSSLs of both families, as shown by the clear clusters formed by the SSSLs (Fig. 6a). Differences between treatments were smaller than differences among the SSSLs. We further identified the existence of several SSSL-specific differentially expressed genes (Fig. 6b, Tables S11b-e and S12b-e). Many more SSSL-specific genes were differentially expressed as a result of exposure to GR24 than those common to all SSSLs of a family, indicating the lack of a strongly conserved transcriptional response to strigolactone, even within an AMF species. Notably, there were more SSSL-specific GR24 responsive genes in dikaryons (75 genes) than among homokaryons (21 genes) (Fig. 6b).

GR24-induced differential gene expression

In both SSSL families, a small number of genes were differentially expressed on exposure to GR24 compared to controls (Fig. 6c). In homokaryons, 2 out of 18 745 transcribed genes were upregulated by GR24. In dikaryons, 7 out of 22 944 transcribed genes were significantly upregulated (Figs. 6c and S6; Tables S11a and S12a). Protein domain analysis revealed functional differences among GR24-upregulated genes in homokaryon and dikaryons (Tables S13a-b). Exposure to GR24 triggered upregulation in homokaryons of an enzyme involved in building fungal cell walls (fungal chitin synthase). The upregulation of enzymes involved in biomolecule synthesis (cytochrome P450) (Cresnar & Petric, 2011) and potential cell-cycle control or signal trafficking regulation (E3 ubiquitin-protein ligase) (Clague & Urbe, 2017) occurred in dikaryon SSSLs. A newly annotated protein (CAB4428400.1), with seven transmembrane helices, was only upregulated in dikaryon SSSLs. S-adenosyl-L-methionine dependent methyltransferase expression was upregulated in all SSSLs in response to GR24, suggesting the role of methylation of biomolecules potentially including DNA, proteins and small-molecule secondary metabolites in the response to strigolactone. Interestingly, there were differences in the number of upregulated genes for this class of methyltransferases between the two lineages. Dikaryon SSSLs showed three upregulated S-adenosyl-L-methionine dependent methyltransferases, while only one was upregulated in homokaryon SSSLs.

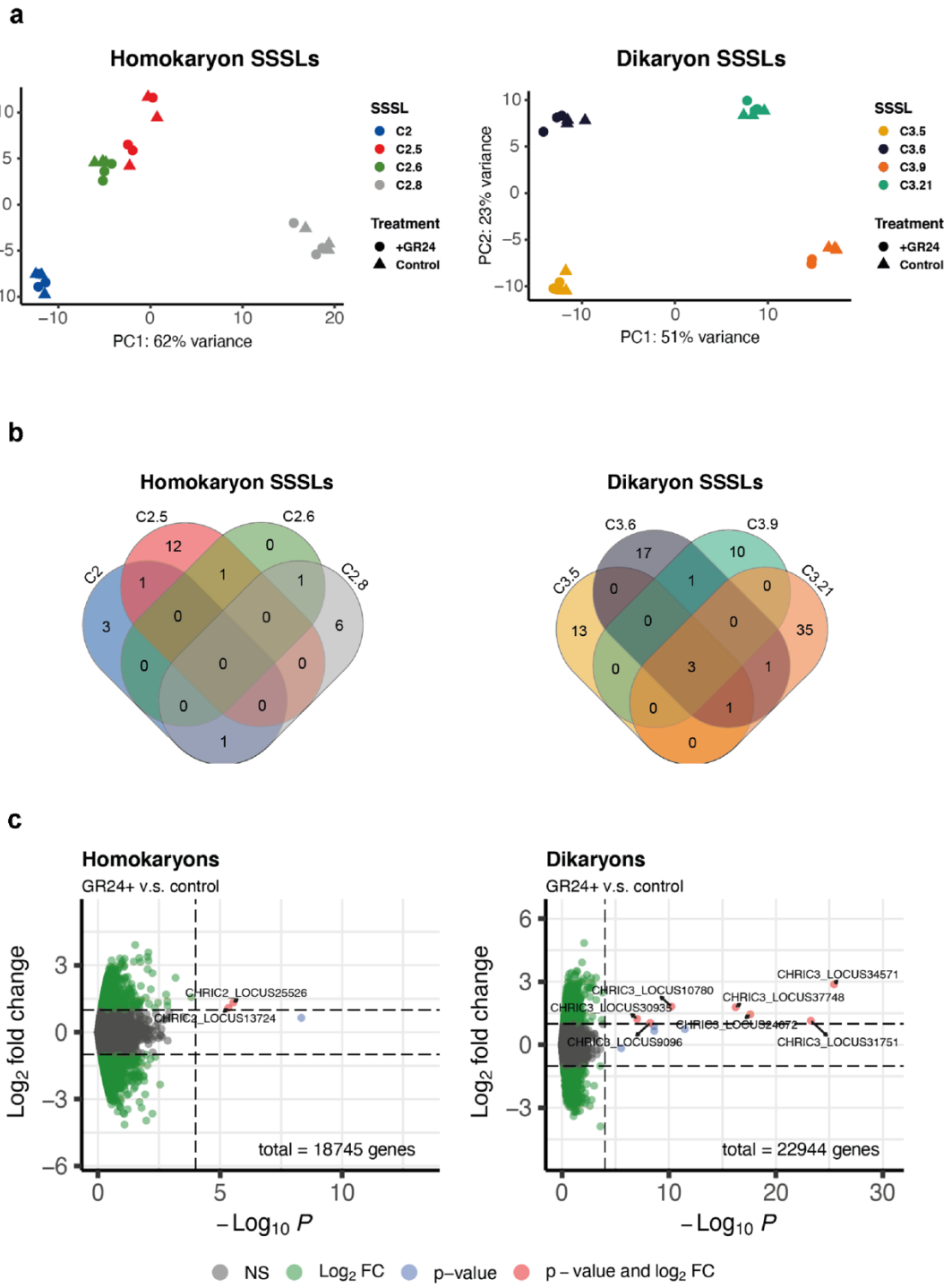


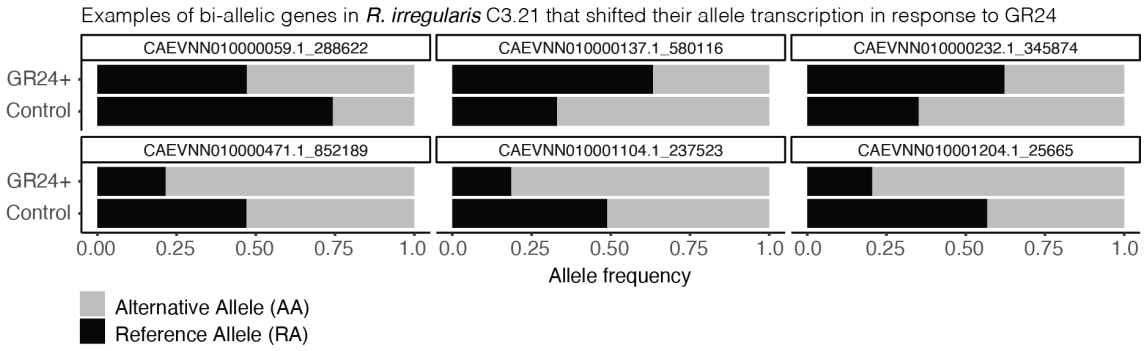
Fig. 6. Transcriptional differences between treatments in homokaryon and dikaryon SSSL families. a) Principal component analysis of the transcriptome data. The first two principal components are displayed with their respective percentages of explained variance in the score plot. b) Venn diagrams displaying the number of differentially expressed genes between the treatments (GR24+ versus Controls) among SSSLs of the two families. c) Volcano plots display the differential gene transcripts between GR24-treated samples and untreated controls for both SSSL families. Genes coloured in red are those considered to be differentially expressed (log2(Fold Change) > 1; FDR < 0.05).

GR24-induced shifting of bi-allelic gene transcription in dikaryons

While GR24 did not affect nucleus genotype proportions, it did affect the proportion expression by the two alleles in some bi-allelic genes (Fig. 7). Six examples of genes that showed GR24-induced opposite shifts in expression of their two alleles are shown in Fig. 7a. This occurred in 11.3% - 18.7% of all transcribed bi-allelic genes and in each of the four dikaryon siblings (Fig. 7b). The proportional change in transcription induced by GR24 of the reference allele of genes ranged from -0.38 (reduction in reference allele transcription) to 0.42 (increase in reference allele transcription), and the median shift ranged from -0.07 to 0.05, depending on the SSSL (Fig. 7b).

We conducted a gene set enrichment analysis to discover whether the bi-allelic genes affected by GR24 in a given SSSL shared common functions. The gene sets showing a shift in allele frequency towards reference or alternative allele were separated and analysed separately. In all SSSLs, there were GR24 affected gene sets representing a given function or process, where the response of the reference allele was in the same direction for the whole gene set representing a function or process (Table S14). This strongly indicates that shifts in bi-allelic gene expression in response to GR24 are not random. A number of functions represented by enriched gene sets, where the alternative allele was expressed more in response to GR24, were shared between SSSLs C3.5 and C3.6 (Table S14a and c). These gene sets represented protein refolding (GO:0042026, GO:0005515), ATP metabolism related processes (GO:0017111, GO:0051082), spliceosome formation (KEGG:map03040) and endocytosis (KEGG:map04144). GR24 induced an increase in reference allele expression for ribosome biogenesis (GO:0042254) and carbohydrate related metabolism (GO:0008152, GO:0071704, GO:1901135 or GO:0005975) in C3.5, C3.6 and C3.21 (Table S14b, d and h). In C3.9, GR24 stimulated reference allele expression for septin ring formation (GO:0005940) and cytoskeleton-dependent cytokinesis (GO:0061640) (Table S14f). Together, this means that for specific cellular functions, collectively in a whole gene set for that function, one allele is induced or repressed in response to GR24. However, it should be noted that a large number of gene sets representing functions that were significantly affected by GR24 occurred that were not common among SSSLs (Tables S14a-h).

a



b

	C3.5		C3.6		C3.9		C3.21	
Dominant nucleus genotype	MAT-1 type		MAT-1 type		MAT-2 type		MAT-2 Type	
# of bi-allelic genes	1489		1485		1488		1467	
# of bi-allelic genes afected by GR24 (%)	279 (18.74%)		167 (11.25%)		185 (12.43%)		173 (11.79%)	
Degree of frequency change of the reference allele induced by GR24	**Range**	**Median**	**Range**	**Median**	**Range**	**Median**	**Range**	**Median**
	[-0,35 , 0.23]	-0.07	[-0,34 , 0.25]	-0.04	[-0,32 , 0.26]	0.03	[-0,38 , 0.42]	0.05
# of shifts toward each allele type (Aleternatiive Allele (AA) & Reference Allele (RA))	**AA**	**RA**	**AA**	**RA**	**AA**	**RA**	**AA**	**RA**
	178	101	97	70	85	100	78	95
AA/AR	1.76		1.39		0.85		0.82	

Fig. 7. GR24-induced shifting of reference allele expression in bi-allelic genes in dikaryon SSSLs. a) Six examples of bi-allelic genes in the SSSL C3.21 where allele expression frequency shifted in response to GR24. b) A summary table of allele frequency shifts of bi-allelic genes induced by GR24 among SSSLs. The number of genes affected by GR24 were based on statistically significant changes (Welch's t-test, $p < 0.05$ followed by following Benjamini-Hochberg correction for multiple testing) in reference allele expression between GR24-treated samples and untreated controls. The lines with MAT1 nucleus genotype dominancy (C3.5 and C3.6) showed a greater number of number of genes shifting towards the alternative allele. An opposite pattern of allele expression shifting is observed in the lines with MAT2 nucleus genotype dominancy (C3.9 and C3.21).

Discussion

The pre-symbiotic stage in AMF is a critical moment in which the fungus has a limited time to form a symbiosis. Here, we observed that the strigolactone GR24 had little effect on germination of homokaryon SSSLs and only on germination of some dikaryon SSSLs. Strigolactone effects on the metabolome were more pronounced among dikaryon SSSLs than among homokaryons. Strigolactone-induced metabolites varied considerably among SSSLs, with little evidence of a commonly conserved metabolomic response across all fungal lines. Contrary to the hypothesis, proportions of the two nucleus genotypes were not affected by exposure to strigolactone in dikaryons. However, gene expression in dikaryon SSSLs was affected by strigolactone, and most strigolactone-responsive genes were SSSL-specific. While nucleus genotype proportions were not affected by GR24, we found a GR24-induced shift in the proportion of gene expression of the two alleles in bi-allelic genes. This occurred in gene sets representing specific cellular functions where many genes of the same set responded in the same way to GR24. We discuss these findings below.

Germination differences among SSSLs are more likely linked to genetic factors than epigenetic variation

We hypothesized that trait variation among *R. irregularis* SSSLs would arise as a manifestation of epigenetic variability among SSSLs or because of unequal nucleus genotype ratios. Our results showed that no trait differences occurred among homokaryon SSSLs in response to GR24. On the contrary, we found the evidence for variation among SSSLs of dikaryons in response to GR24.

We confirmed that proportions of nucleus genotypes varied among dikaryon siblings. The results are consistent with reports that C3.9 and C3.21 displayed unequal nucleus genotype proportions (Masclaux *et al.*, 2018; Robbins *et al.*, 2021). However, qPCR also revealed an unequal proportion of nucleus genotypes in C3.5 and C3.6. This was not consistent with previous studies (Masclaux *et al.*, 2018; Robbins *et al.*, 2021).

Given that differences in nuclear genotype proportions were the only known genetic differences among these clonally-produced sibling dikaryons, it is likely that observed differences among dikaryon SSSLs are most likely linked to the co-existence of 2 different nucleus genotypes than to a potential epigenetic reprograming induced by GR24.

Previously, Serghi *et al.* (2021) claimed very different life-history strategies between homokaryons and dikaryons, with homokaryons exhibiting higher spore germination rates and lower germination times compared to dikaryons. Interestingly, the results in the present study showed the opposite. Indeed, while most dikaryons spores had germinated after 7 days (> 80% of the spores), it took between 14 and 21 days for homokaryons to reach similar germination rates (Fig. S3).

Lack of conserved AMF trait, metabolic and transcriptional GR24 responses

Plants have a highly conserved "toolbox" of genes allowing mycorrhizal symbiosis establishment (Schmitz & Harrison, 2014). Plant release of strigolactones is thought to be the initial stage in the process initiating symbiosis (Besserer *et al.*, 2006; Schmitz & Harrison, 2014; Lanfranco *et al.*, 2018). Because the AMF response to strigolactone is assumed to be a critical step towards the formation of the symbiosis, we expected this response to be conserved such that germination of all SSSLs would be affected by GR24 and that the metabolomic and transcriptomic response would be similar. We expected that any differences among SSSLs in response to GR24 would be mostly quantitative (the abundance of a given metabolite; or the expression levels of a set of genes). Remarkably, this is not what we observed.

The untargeted metabolomic profiling does not inform on the identity of metabolites involved in the process of spore germination. Here we provide a first glimpse on the variability of exudates within the model AMF species *R. irregularis*. We observed SSSL-specific differences in which metabolites were produced in response to GR24, especially among dikaryon SSSLs (Fig. 4, Fig. 5). This surprising lack of conserved metabolic response among SSSLs was largely mirrored in the transcriptomes. A remarkably small number of common genes were differentially expressed in response to GR24 (Fig. 6). Most notably, GR24 affected the expression of a much larger number of genes that were SSSL-specific and this was more pronounced in dikaryons than homokaryons. This leads us to conclude that expression of some GR24-responsive genes, observed in some homokaryon lines, may potentially be under epigenetic control. However, expression of a notably larger number of GR24-responsive genes that were SSSL-specific in the dikaryons suggests that additional quantitative genetic effects play a role in responsiveness to strigolactones. Nucleus genotype ratios could potentially influence the GR24-responses in dikaryons.

The lack of a conserved metabolomic and transcriptomic response to GR24 in *R. irregularis* would explain why we observed the absence of significant differential expression in the SSSLs of C2 and C3 among the set of 19 GR24 responsive genes previously described by Tsuzuki *et al.* (2016) in DAOM197198 (Tables S5 and S6).

Relationship between nucleus genotype proportions and transcription of bi-allelic genes

Previously, we suggested that allele frequency of transcription of bi-allelic genes was influenced by nucleus genotype proportions because C3.9 and C3.21 showed unequal nucleus genotype proportions and an unequal transcription of two alleles of bi-allelic genes (Robbins *et al.*, 2021). To support this, an equal transcription of the two alleles at bi-allelic genes was observed in C3.5 and C3.6 that appeared to have equal proportions of two nucleus genotypes (Robbins *et al.*, 2021). In this study, the results were consistent in C3.9 and C3.21 with the study by Robbins *et al.* (2021). However, unequal nucleus proportions were also observed in C3.5 and C3.6, while gene transcription of bi-allelic genes appeared to be equal for both alleles.

Nucleus genotype proportions and transcriptional responses to strigolactone

We hypothesized that exposure to GR24 would alter nucleus genotype ratios in dikaryon SSSLs. While dikaryon SSSLs differed in their nucleus genotype ratios, the ratios were not affected by GR24 (Fig. 5a). Overall, allele frequency distributions of gene expression in bi-allelic genes differed among SSSLs but the average response of each SSSL was unaffected by GR24. Our findings suggest that nucleus genotype ratios in dikaryons are stable to some environmental perturbations, at least in pre-symbiotic stage.

Most interestingly, we found that dikaryons adjust the ratio of expression of two alleles of bi-allelic genes in response to GR24, in the absence of any GR24-induced change nucleus genotype proportions. Similar GR24-induced changes in which allele of a bi-allelic gene was expressed the most were commonly observed in C3.5 and C3.6 and C3.9. Many of the genes affected were members of gene sets related to the same functions. One particularly notable case concerns the bi-allelic gene set related to the spliceosome. The spliceosome delivers the fundamental cellular mechanism known as alternative splicing. So while GR24 did not affect the expression of a very large number of genes, it is likely that GR24-induced alternative splicing could have affected protein isoform formation by forming premature stop codons rather than affecting gene expression levels (Reddy, 2007). This would explain the high degree of differences in exudate profiles without large changes in differentially expressed gene in dikaryon lines. Moreover, in plants, it is already known that external stresses can induce alternative splicing; a form of regulation that is crucial in polyploids (Zhou *et al.*, 2011). The observed favouring of specific allelic expression for the spliceosome suggests that differential gene expression (either quantitative or qualitative) in response to an environmental cue is not the sole mechanism to regulate cellular functions in AMF. Indeed, post-transcriptional regulation could be a more efficient response to an environmental change than the adjustment of nucleus genotype proportions. The current short-read transcriptome analysis used here is not optimal for detecting alternative splicing. We suggest that long-read RNA sequencing should be considered to understand these complex gene regulation mechanisms in AMF in a changing environment.

We also found a common preference in allelic expression of genes controlling ribosome biogenesis and carbohydrate-related metabolism among some SSSLs in response to GR24. Strigolactone signals the existence of a nearby host. Our results suggest that the signal stimulates AMF to alter the conversion of stored energy to metabolism and alter gene expression towards a larger ribosome population for the upcoming interaction with host. It should be emphasized that an AMF spore has limited amount of stored energy to perform cellular functions, but that after forming symbiosis the energy source is no longer limited. Thus, during the symbiotic phase, it may be advantageous to re-adjust the nucleus genotype proportions in response to an environmental change. However, this may be too slow or too costly in the pre-symbiotic phase and a more successful strategy could be that the fungus rapidly shifts expression of different alleles of bi-allelic genes.

Conclusions

Lastly, we want to highlight the ecological and evolutionary meaning of our findings. There is some debate about whether AMF are sexual or asexual. While this study does not address that debate, it is notable that in our experiments, where individual SSSLs are grown *in vitro*, there is no possibility for sexual reproduction. AMF dikaryons inherit same qualitative genomic pool with quantitative differences among SSSLs in how that information is packaged in their asexual spores. Here, we show that clonal reproduction of AMF can still allow the production of offspring with diverse fitness. Even during clonal growth, differences in proportions of nucleus genotypes, and the ability independently of nucleus genotype proportions, to regulate gene expression in response to critical environmental cues give clonally growing dikaryon AMF considerable plasticity above that normally expected for a clonal organism.

Data Availability

Raw RNA-sequencing data are accessible through the European Nucleotide Archive under the accession number: PRJEB52540. All additional data are provided as supplementary files.

Acknowledgements

We would like to thank Eric Risse and Consolee Aletti for assistance. We thank the Lausanne Genomics Technology Facility (GTF) for sequencing the RNA libraries. This research was funded by the Swiss National Science Foundation (project no. 310030B_182826).

Author Contributions

JCC, SJL and CR conceived the study. JCC and SJL performed data analysis. JCC, SJL, GG and CR performed experimental work. JCC, SJL and IRS contributed to writing and editing the manuscript. SJL and JCC contributed equally to this work.

References

Anders S, Pyl PT and Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2): 166-169.

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Angelard C, Colard A, Niculita-Hirzel H, et al. (2010) Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Curr Biol* 20(13): 1216-1221.

Angelard C, Tanner CJ, Fontanillas P, et al. (2014) Rapid genotypic change and plasticity in arbuscular mycorrhizal fungi is caused by a host shift and enhanced by segregation. *ISME J* 8(2): 284-294.

Besserer A, Puech-Pages V, Kiefer P, et al. (2006) Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria. *PLoS Biol* 4(7): e226.

Bolger AM, Lohse M and Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114-2120.

Brundrett MC and Tedersoo L (2018) Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytol* 220(4): 1108-1115.

Ceballos I, Mateus ID, Peña R, et al. (2019) Using variation in arbuscular mycorrhizal fungi to drive the productivity of the food security crop cassava. *bioRxiv [Preprint]*. DOI: https://doi.org/10.1101/830547.

Chaturvedi A, Cruz Corella J, Robbins C, et al. (2021) The methylome of the model arbuscular mycorrhizal fungus, Rhizophagus irregularis, shares characteristics with early diverging fungi and Dikarya. *Commun Biol* 4(1): 901.

Clague MJ and Urbe S (2017) Integration of cellular ubiquitin and membrane traffic systems: focus on deubiquitylases. *FEBS J* 284(12): 1753-1766.

Cresnar B and Petric S (2011) Cytochrome P450 enzymes in the fungal kingdom. *Biochim Biophys Acta* 1814(1): 29-35.

Dobin A, Davis CA, Schlesinger F, et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1): 15-21.

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

Keymer A, Pimprikar P, Wewer V, et al. (2017) Lipid transfer from plants to arbuscular mycorrhiza fungi. *Elife* 6.

Koch AM, Antunes PM and Klironomos JN (2012) Diversity effects on productivity are stronger within than between trophic groups in the arbuscular mycorrhizal symbiosis. *PLoS One* 7(5): e36950.

Kokkoris V, Chagnon PL, Yildirir G, et al. (2021) Host identity influences nuclear dynamics in arbuscular mycorrhizal fungi. *Curr Biol* 31(7): 1531-1538 e1536.

Lanfranco L, Fiorilli V, Venice F, et al. (2018) Strigolactones cross the kingdoms: plants, fungi, and bacteria in the arbuscular mycorrhizal symbiosis. *J Exp Bot* 69(9): 2175-2188.

Love MI, Huber W and Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12): 550.

Martinez Arbizu P (2020) pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4.

Masclaux FG, Wyss T, Mateus-Gonzalez ID, et al. (2018) Variation in allele frequencies at the bg112 locus reveals unequal inheritance of nuclei in a dikaryotic isolate of the fungus Rhizophagus irregularis. *Mycorrhiza* 28(4): 369-377.

McGinnis S and Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 32(Web Server issue): W20-25.

Nuccio EE, Hodge A, Pett-Ridge J, et al. (2013) An arbuscular mycorrhizal fungus significantly modifies the soil bacterial community and nitrogen cycling during litter decomposition. *Environ Microbiol* 15(6): 1870-1881.

Oksanen J, Kindt R, Legendre P, et al. (2020) vegan: Community Ecology Package.

Pena Venegas RA, Lee SJ, Thuita M, et al. (2021) The Phosphate Inhibition Paradigm: Host and Fungal Genotypes Determine Arbuscular Mycorrhizal Fungal Colonization and Responsiveness to Inoculation in Cassava With Increasing Phosphorus Supply. *Front Plant Sci* 12: 693037.

R Core Team (2020) A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol* 58: 267-294.

Robbins C, Cruz Corella J, Aletti C, et al. (2021) Generation of unequal nuclear genotype proportions in Rhizophagus irregularis progeny causes allelic imbalance in gene transcription. *New Phytol* 231(5): 1984-2001.

Ropars J, Toro KS, Noel J, et al. (2016) Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nat Microbiol* 1(6): 16033.

Sanders IR and Croll D (2010) Arbuscular mycorrhiza: the challenge to understand the genetics of the fungal partner. *Annu Rev Genet* 44: 271-292.

Savary R, Masclaux FG, Wyss T, et al. (2018) A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus Rhizophagus irregularis. *ISME J* 12(1): 17-30.

Schmitz AM and Harrison MJ (2014) Signaling events during initiation of arbuscular mycorrhizal symbiosis. *J Integr Plant Biol* 56(3): 250-261.

Serghi EU, Kokkoris V, Cornell C, et al. (2021) Homo- and Dikaryons of the Arbuscular Mycorrhizal Fungus Rhizophagus irregularis Differ in Life History Strategy. *Front Plant Sci* 12: 715377.

Smith S and Read D (2008) Mycorrhizal Symbiosis. 3rd Edition ed.: Academic Press.

Tsuzuki S, Handa Y, Takeda N, et al. (2016) Strigolactone-Induced Putative Secreted Protein 1 Is Required for the Establishment of Symbiosis by the Arbuscular Mycorrhizal Fungus Rhizophagus irregularis. *Mol Plant Microbe Interact* 29(4): 277-286.

van der Heijden MG, Klironomos JN, Ursic M, et al. (1998) Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* 396: 69-72.

van der Heijden MG, Wiemken A and Sanders IR (2003) Different arbuscular mycorrhizal fungi alter coexistence and resource distribution between co-occurring plant. *New Phytol* 157(3): 569-578.

von Mering C, Huynen M, Jaeggi D, et al. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 31(1): 258-261.

Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.

Wyss T, Masclaux FG, Rosikiewicz P, et al. (2016) Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus Rhizophagus irregularis. *ISME J* 10(10): 2514-2526.

Zhou R, Moshgabadi N and Adams KL (2011) Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proc Natl Acad Sci U S A* 108(38): 16122-16127.

Supplementary Information

Supplementary figures

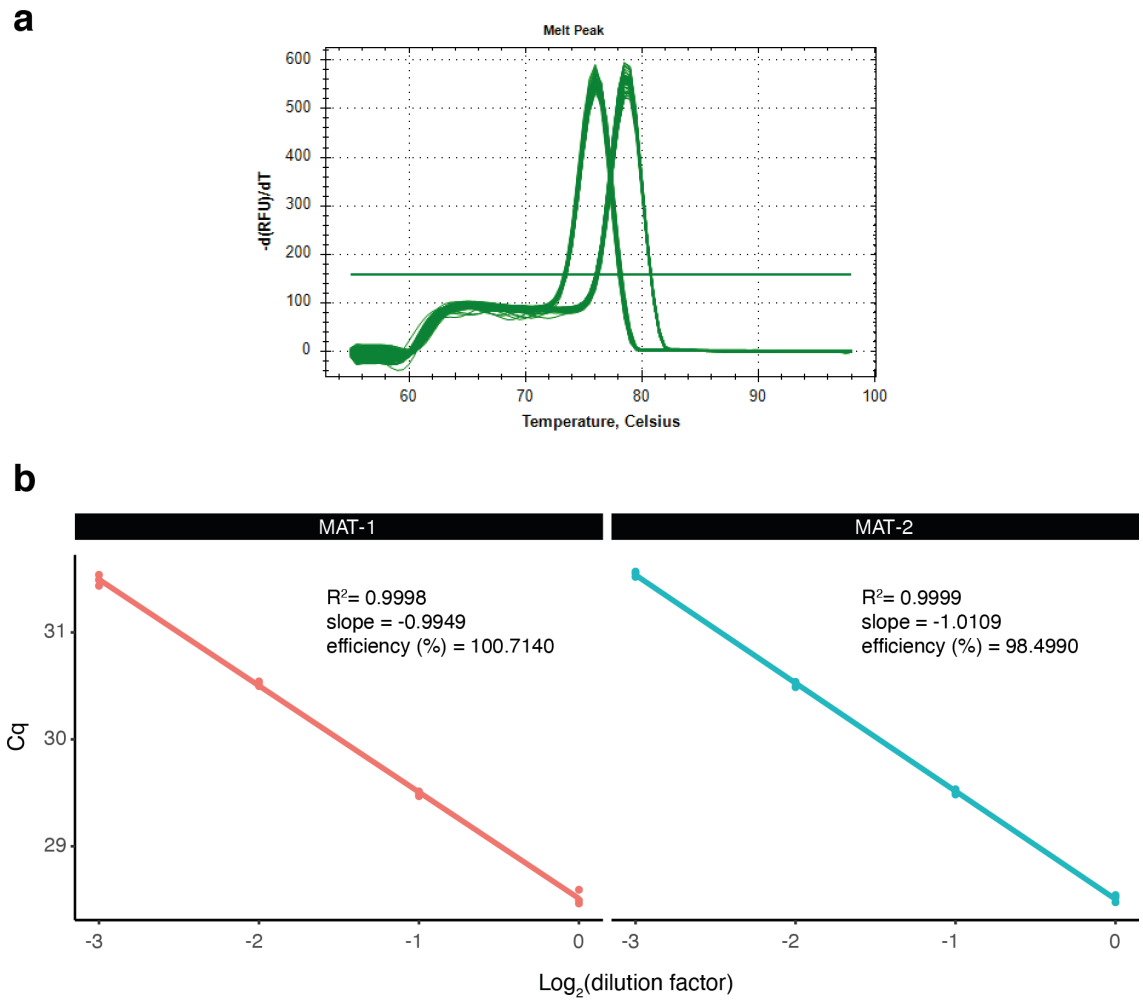


Fig S1. Tests of qPCR primer specificity and efficiency used in the qPCR assays to determine the relative abundance of each nucleus genotype in dikaryon SSSLs. a) The PCR product of each primer set (used to amplify MAT-1 and MAT-2 alleles) showed no off-target amplification. -d(RFU)/dT values represent the rate of change in the relative fluorescence units (RFU) with time (T). b) Efficiency of the two sets of primers. The efficiency of the primers, the slope of the regression line and the regression coefficient (R^2) are reported for each set. Cq values indicate the quantification cycle.

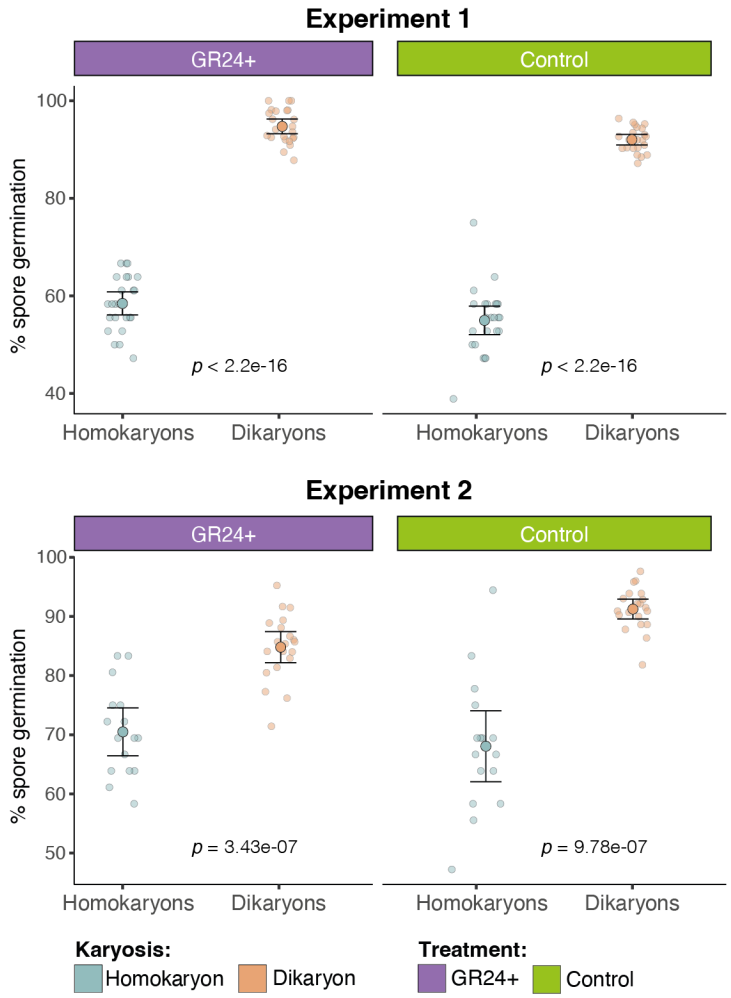


Fig S2. Spore germination differences between the two SSSLs families in the two experiments. The mean percentage of spore germination of each R. irregularis SSSLs family is shown for GR24 treated (GR24+) and untreated (Control) samples 7 days after inoculation. Individual replicate units are represented by semi-transparent data points. Solid dots and error bars represent the mean percentages of germination and the 95% confidence interval. Probability (P) of the test for a difference between homokaryon and dikaryon spore gemination % is shown. In experiment 1, homokaryon SSSLs had a significantly lower mean of spore germination than dikaryon SSSLs when treated with GR24 (M=58.45, SD=5.58) and (M=94.74, SD=3.50), (t = -26.8, df = 38.93, p < 2.2e-16); and in the untreated controls (M=54.98, SD=6.90) and (M=92.03, SD=2.49), (t = -24.68, df = 29.09, p < 2.2e-16). In experiment 2, homokaryon SSSLs had slightly higher means of spore germination compared to experiment 1, but still significantly lower than dikaryon SSSLs when treated with GR24 (M=68.05, SD=11.30) and (M=91.25, SD=3.60), (t = -7.93, df = 17.47, p = 3.43e-07); and in the untreated controls (M=70.49, SD=7.58) and (M=84.81, SD=5.62), (t = -6.29, df = 26.98, p= 9.78e-07).

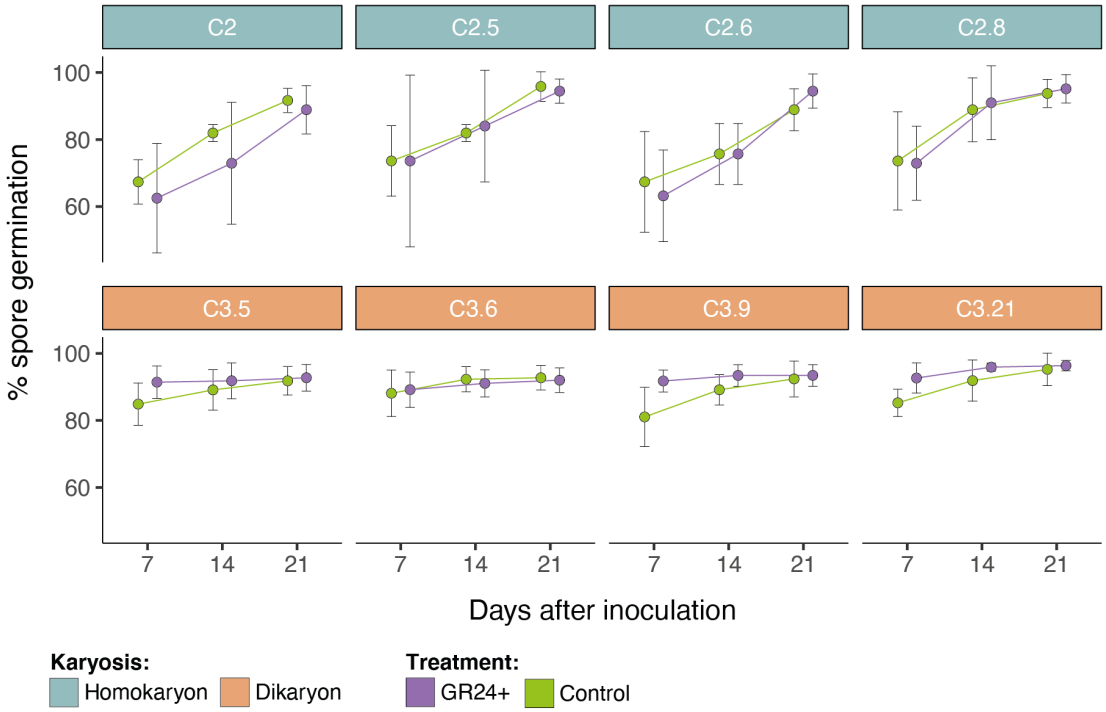


Fig S3. Spore germination (%) at three different time points in experiment 2. The mean percentage of germinated spores of each SSSL is shown for GR24 treated (GR24+) and untreated (Control) samples 7,14 and 21 days after being put onto the solid medium. The error bars represent the 95% interval of confidence of the mean. Data points of GR24+ and Controls have been shifted horizontally to avoid overlapping of points on the x-axis and improve the readability of the results. However, spore germination was measured at exactly the same time points in both treatments.

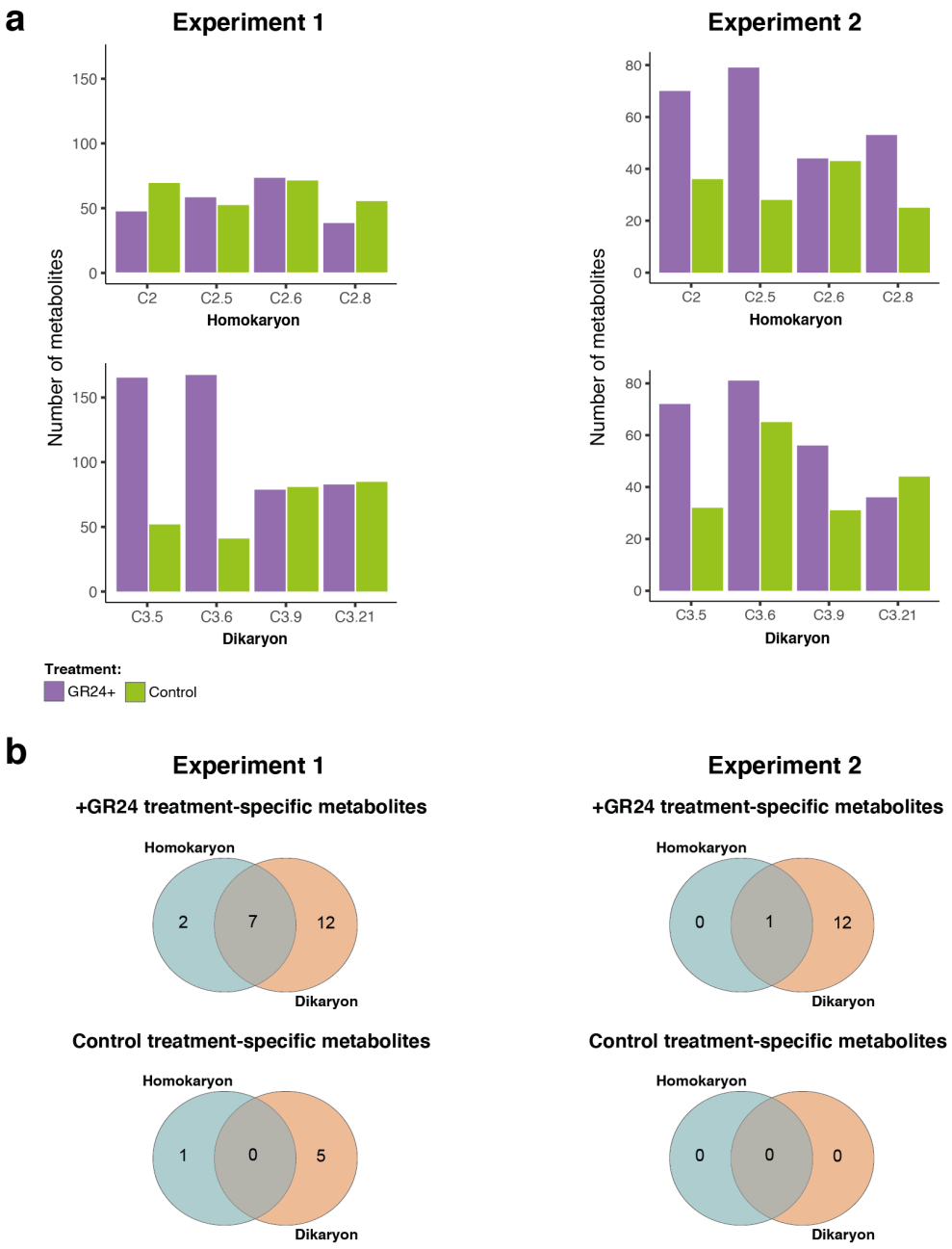


Fig S4. Treatment-specific metabolite counts among SSSLs in experiments 1 and 2. a) Barplots displaying the number of metabolites specific to each experimental treatment (GR24+ / Controls) of each SSSL. b) Venn diagrams displaying the number of metabolites shared among SSSLs of the same family and between the two SSSL families.

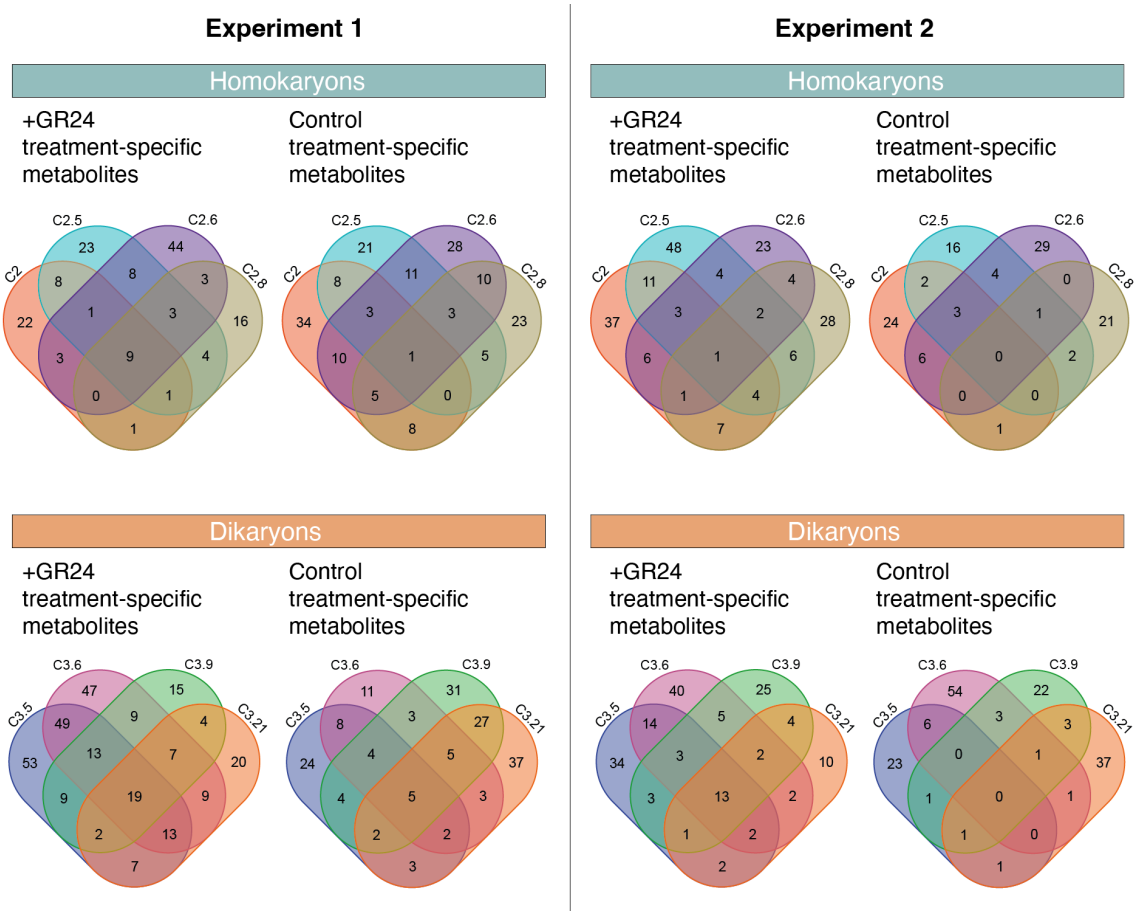


Fig S5. Treatment-specific metabolite counts shared among SSSLs in experiments 1 and 2**.** Venn diagrams displaying the number of common and isolate-specific metabolites detected using LC-MS.

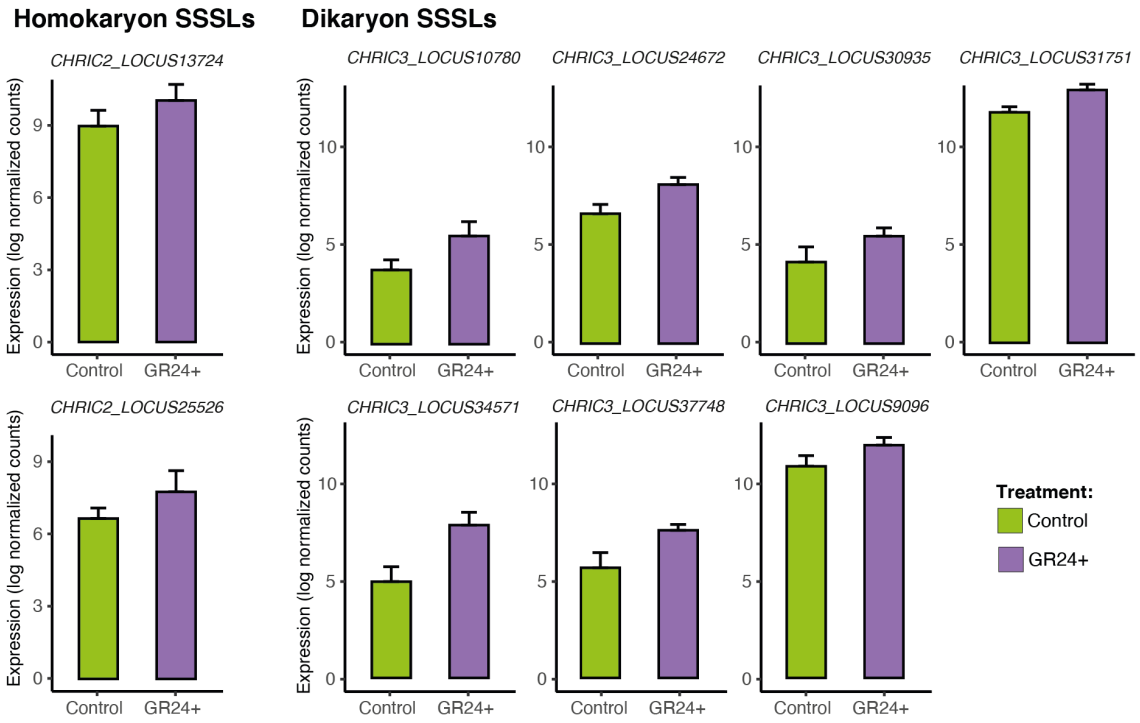


Fig S6. Differential gene expression induced by GR24 in homokaryon and dikaryon SSSLs. GR24 induced significantly different expression in 2 genes among the homokaryon SSSLs (left side panel), and 7 genes among the dikaryon SSSLs (right side panel). Barplots show the mean expression values (log normalized counts) and the error bars represent the standard deviation.

Supplementary tables

Before the manuscript publication, all the supplementary tables can be found in a Supplementary MS Excel File in the URL below:

https://github.com/Jquimcrz/SI_Tables_Chapter4_PhD/blob/main/Supplementary%20Tables.xlsx

Chapter 5

General discussion

Summary of the findings

In **chapter 2**, I found that *R. irregularis* isolates displayed DNA methylation profiles that set them apart from other fungal taxa. More specifically, I found that 6mA in *R. irregularis* displayed signatures of a functional epigenetic mark potentially regulating important processes for the AM symbiosis, despite its relatively low abundance in this species. In **chapter 3**, I found that dikaryon SSSLs displayed quantitative genetic variation in the form of variable nucleus genotype ratios, which in turn caused an imbalanced allele transcription during gene expression. The most abundant nucleus genotype was most likely dominant in terms of transcriptional output. Furthermore, I also found remarkable evidence of mono-allelic expression of several bi-allelic genes that displayed high SNPs densities, which is compatible with the existence of an additional layer of epigenetic regulation. Lastly, in **chapter 4**, I found that germination differences among dikaryon *R. irregularis* SSSLs were most likely related to the underlying existing differences in nucleus genotype ratios; although further profiling of additional molecular data revealed a surprising lack of a conserved metabolomic and transcriptomic response among *R. irregularis* siblings to a strigolactone, even among homokaryon SSSLs. Altogether, the results gathered in this thesis show that both, epigenetics and nuclear ratios in dikaryons, can potentially allow variation in gene expression and, ultimately, contribute to phenotypic trait variation in AMF. This is important given that AMF are largely clonal microorganisms. In the next sections, I will discuss the most notable consequences of my findings.

Towards a better understanding of the epigenetic regulation in AMF

The epigenetic regulation of a genome can contribute considerably in determining which genes are expressed and under which circumstances. In eukaryotes, the chemical modification of the DNA sequence and of histones influence the rate at which transcription factors and other DNA binding proteins interact with gene regulatory sequences (promoters, enhancers, silencers, insulators, etc.) (Allis and Jenuwein, 2016). The dynamic nature of epigenetic regulation is key to many cellular processes that are time and context dependent, and that start with a specific stimulus that trigger a wider signalling cascade. Studying these cellular processes and their regulation is critical to understand how organisms develop and respond to the environment. Since AMF are thought to be reproducing largely by asexual reproduction, the ability to modify gene expression without genetic change would potentially provide increased plasticity of AMF in fluctuating environments.

The presence of two functional DNA methylation epigenetic marks (6mA and 5mC) in the *R. irregularis* genome highlights that epigenetic regulation is most likely playing an important role also in AMF, a group of organisms that has thrived in a large range of environments and associating with a wide range of plant species. Both of these epigenetic marks had been studied in closely related fungal species, including some EDF and some dikarya (Mondo et al., 2017; Bewick et al., 2019). Notably, these two major fungal clades evolved quite different epigenetic regulation strategies. EDF retained significantly higher percentages of 6mA methylation than dikarya, and displayed evidence of that

being a functionally active epigenetic mark by promoting gene expression. Instead, EDF displayed very low levels of 5mC methylation compared to Dikarya. The fact that AMF have a distinct methylation profile compared to other EDF and to dikarya fungi is perhaps an indication that their long-term obligate symbiotic lifestyle has shaped their epigenetic landscape to accommodate their particular biology. There is extensive evidence that symbioses represent a source of selectable epigenetic variation (Gilbert et al., 2010; Asgari, 2014), and it is likely that such an ancient and close interaction between AMF and plants has not only resulted in a process co-evolution of their genomes, but also of their epigenomes.

We showed that 5mC is an abundant epigenetic mark in the *R. irregularis* genome, which is not a feature shared by the other EDF species. However, the lack of resolution in the Pacific Biosciences data precluded a more detailed study of the role of 5mC in *R. irregularis*. A recent study shed some light on this topic by showing that 5mC is most likely acting as a regulator of transposable element (TE) activity (Dallaire et al., 2021). The *R. irregularis* genome is, indeed, very rich in TEs and it is thought that they greatly contributed to its evolution, although this subject still remains unclear in this species.

In addition to 5mC, we found that 6mA was also a key epigenetic modification in this AMF species. Although a much smaller set of genes was regulated by 6mA in *R. irregularis* than in other EDF, it was enriched in genes thought to be involved in core processes that are important for the symbiosis, such as nutrient transport and metabolism.

A main focus of this PhD thesis was the variability in AMF possible by a homokaryotic or dikaryotic state. Most notably, even though *R. irregularis* isolates displayed a largely conserved 6mA methylome, we still found some variability in the gene sets regulated by 6mA between two homokaryon isolates C2 and C5 that were genetically indistinguishable. Given the lack of any observable genetic variation between C2 and C5, it is likely that a form epigenetic variation is responsible for their phenotypic differences, although this would probably not be restricted to adenine methylation. Understanding how trait variation arises among homokaryon clonal isolates is a critical step to advance in the agronomic application of AMF, because studies such as Ceballos *et al.* (2019) and Peña *et al.* (2020) provided evidence of differential effects exerted by homokaryon SSSLs on cassava in randomized field trials conducted in Colombia and in Kenya (Ceballos et al., 2019; Peña et al., 2020). These were observed both in plant growth and other plant physiological traits. The variability in plant growth observed among plants inoculated with homokaryon SSSLs was comparable to those inoculated with dikaryon SSSLs, or between other genetically different AMF isolates, and in some cases even higher. While other covariates such as the soil microbiome or the soil physicochemical properties could certainly influence the cassava traits assessed in these field studies, AMF epigenetics most likely represents an important source of plant trait variation by, either directly affecting the regulation of nutrient exchange, or indirectly affecting the ecological interactions of whole system, often referred as a holobiont (Vandenkoornhuyse et al., 2015).

It is important to mention that, in this thesis, we focused on the characterization of 6mA in the genome of *R. irregularis*. This, clearly, only represents one form of DNA methylation in only one AMF species. Many other well-known forms of epigenetic regulation (histone modification marks such as H3K27ac; RNA modifications; transcription factors binding, etc.) should be, therefore, considered in future experiments. For instance, one could attempt to understand the genome-wide differences in chromatin accessibility in several AMF isolates when exposed to different hosts, grown under different nutrient conditions or at different developmental stages by performing an Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq). In the case of dikaryon AMF, a combination of single-nucleus ATAC-seq (snATAC-seq) and single-nucleus RNA-seq could be a very interesting and suitable approach to better understand the role of the two different nucleus genotypes populations and their actual contribution to the whole transcriptome. These are well-stablished techniques in many other biological systems that might be challenging to optimize for their application in AMF, but their results would greatly inform us about this subject. Additionally, Chromatin Immunoprecipitation sequencing (ChIP-seq) experiments could certainly help to elucidate which are the epigenetic marks mediating the chromatin accessibility changes in AMF. Lastly, it is also important to expand AMF research beyond the model species *R. irregularis* to generalize the findings to the rest of the Glomeromycotina.

Origin and regulation of variable nuclear ratios in dikaryon AMF

Several studies over the past ten years shed light on the possible origin, regulation and hypothetical function of the variable nuclear ratios in AMF. Angelard *et al.* (2010) and Ehinger *et al.* (2012) first showed that sibling lines, originated from single spores of the same parent AMF isolate, displayed significantly different phenotypes (Angelard et al., 2010; Ehinger et al., 2012). These two studies suggested that important plant and fungal traits could be strongly influenced by subtle differences in the AMF genetic makeup. These phenotypic differences were linked to changes in allele frequencies in several nucleus-specific loci among AMF SSSLs, although at the time, it was unclear how such quantitative variation originated. The partial segregation of nuclear genotypes during spore formation was considered the most likely explanation, given that changes in allele frequencies were detected in loci likely partitioned among different nucleus genotypes. Later, Angelard *et al.* (2014) showed that different plant hosts induced changes in the allele frequencies of the nucleus-specific bg112 locus among several *R. irregularis* SSSLs (Angelard et al., 2014). The results of Angelard *et al.* (2014) suggested that nuclear ratios in AMF were likely responsive to the environment (i.e. the plant host during the symbiotic stage); and likely related to the AMF enormous adaptability. However, despite the increasing evidence that nuclear ratios varied among AMF SSSLs, it remained unknown how such ratios originated and whether they were actively regulated. Masclaux *et al.* (2018) showed that dikaryon SSSLs harboured different nuclear ratios using more precise molecular methods than Angelard *et al.* (2010) and Ehinger *et al.* (2012) and suggested a stochastic origin of these nuclear ratios, by the random segregation of nuclear genotypes during spore formation (Masclaux et al., 2018).

Kokkoris *et al.* (2021) subsequently showed that dikaryon spores of AMF had significantly different nuclear ratios when exposed to different plant hosts, similarly to what was shown by Angelard *et al.* (2014) (Kokkoris et al., 2021). Remarkably, in chapter 3, we showed that the same SSSLs analysed by Masclaux *et al.* (2018) and originally produced by Angelard *et al.* (2010) still displayed variable nuclear ratios after several years of homogenous culturing with the same plant host; and that the ratios had a direct effect on the transcriptome of the SSSLs (Robbins et al., 2021). These results were less supportive of the hypothesis that nuclear ratios changed in response to the environment, although this hypothesis was not specifically tested in our study.

All these independent pieces of evidence emphasize the importance of studying (1) how different nuclear genotypes ratios originate among dikaryons, (2) how they are regulated, maintained in stable proportions or shift in different environments, and ultimately (3) how they influence AMF phenotypes.

Kokkoris *et al.* (2021) claimed that dikaryon AMF shifted their nuclear genotype proportions according to the identity of the plant host in a non-stochastic manner, although they did not give any evidence for such a directed regulation. Instead, they hypothesized a system that would regulate the proliferation of the most appropriate nuclear genotype; a mechanism that could potentially involve the para-synchronous mitotic division of the two nuclear genotypes. Several multinucleated fungal species, such as *Neurospora crassa* or *Ashbya gossypii*, undergo asynchronous or para-synchronous mitotic divisions, although these have never been observed in AMF (Gladfelter, 2006; Gladfelter et al., 2006; Roca et al., 2010). Because Kokkoris *et al.* (2021) and Angelard *et al.* (2010) both independently reported allele frequency changes when AMF dikaryons were exposed to a different plant host, during the symbiotic phase, a cryptic regulatory system of the nuclear ratios influenced by the plant could potentially occur.

A second possibility to explain the origin and maintenance of the unequal nuclear ratios would be that dikaryon AMF isolates produce thousands of spores each containing a stochastic mixture of the two nuclear genotypes, giving rise to a wide range of possible nuclear configurations. This was first suggested by Masclaux *et al.* (2018). By the same principles that regulate natural selection, only those spores harbouring nuclear ratios that could germinate and thrive, with advantageous or neutral phenotypes in a given environment, would then successfully associate with a plant host and the produce mature mycelia. In this second scenario, the original and stochastic nuclear ratios of the AMF spore could be maintained stably over time with the synchronous mitotic division of both nuclear genotypes, and natural selection would determine whether nuclear ratios in the mycelia are kept or shifted with each generation of germinating spores. In such case, SSSLs harbouring different nuclear ratios could maintain these unequal ratios even when growing under the same environmental conditions (e. g. if there is not a strong selective pressure on the fungus, several nuclear ratios could give rise to equally successful fungi). Indeed, this second possibility seems more conformant with the observations made in chapters 3 and 4, and those from Masclaux *et al.* (2018), where several SSSLs consistently differed in their nuclear ratios even if they were always cultured in the same exact conditions and with the same plant host. In addition, in chapter 4, we observed how the strigolactone

(GR24) applied to germinating spores did not significantly alter the nuclear ratios in any of the four dikaryon SSSLs. Although this is only one environmental cue that cannot be taken as a representative of every possible factor, it showed that the nuclear ratios could be intrinsically more related to the "identity" of the SSSLs, rather than responsive to a certain environment. In either case, further research is necessary in order to understand what the actual mechanism is regulating nuclear genotypes proportions in dikaryon AMF.

Genetic and epigenetic basis of trait variation in AMF

The molecular bases that explain trait variation in AMF remain an open question. Some AMF traits are of high agronomic interest, such as those conferring plants with increased drought tolerance, biomass production or resistance to certain pathogens. Because these phenotypic traits cannot be directly assessed from the fungus itself, they are typically inferred from the effects the fungus exerts on the plants. All the experiments conducted using *R. irregularis* SSSLs have proven that very subtle molecular differences among AMF can still yield quite large and significant trait differences, including those on plant growth as shown by the work of Angelard *et al.* (2010), Ceballos *et al.* (2019) and Peña *et al.* (2020). While this approach of linking AMF genotype to plants traits has many potential practical benefits, it also represents a major limitation for a more thorough understanding of the relationship between genotype and phenotype in AMF.

Complex phenotypes such as growth are usually dependent on combinations of several factors (i.e. genetic, epigenetic, environmental, stochastic, etc.). The gap that exists between the molecular biology of the fungus and the macroscopic phenotypic manifestation in the plant is very large and it may appear difficult to reconcile. Therefore, it becomes crucial to assess and determine the actual contribution of each factor to specific phenotypes.

In this thesis, I investigated the potential contribution of epigenetic and a form of quantitative genetic variation in the emergence of trait variation among *R. irregularis* SSSLs, in the pre-symbiotic stage. This was conducted in a highly controlled environment in order to reduce the number of covariates to the minimum. We observed that homokaryon SSSLs did not differ in their percentages of germinating spores when exposed to a strigolactone. In contrast, two out of four dikaryon SSSLs with similarly biased nuclear ratios consistently showed differential responses to the same strigolactone treatment. While these results were supportive that variable nuclear ratios are most likely involved in the germination differences present among SSSLs, it remains impossible to generalize what would happen to other traits. Homokaryons were primarily used to investigate the potential and sole contribution of epigenetics to trait variation. However, even if germination differences are not strictly epigenetically determined, differences among dikaryons could easily arise from the combination of both, epigenetic and quantitative genetic variation. In fact, the surprising finding that SSSLs, neither homokaryon nor dikaryon, did not share a conserved transcriptomic and metabolomic response to such a widespread signalling molecule like a strigolactone only re-emphasized our current lack of understanding of the

mechanisms that permit AMF adaptation to many environments, and of the basis of AMFs perplexing phenotypic plasticity. Thus, disentangling the effects of genetics and epigenetics on AMF phenotypes still requires further experiments.

Experimental designs including both homokaryon and dikaryon SSSLs, like in chapter 4, are appropriate to first establish which of the two factors, quantitative genetics or epigenetics, is more likely to influence variation in a given trait. However, these experiments can rapidly escalate in size, time and cost depending on the number of experimental conditions, AMF siblings and biological replicates. Thus, perhaps a more reasonable approach would be to focus on one single aspect – epigenetics variability among homokaryons, i.e. - and in a much more reduced set of AMF isolates. Sequencing experiments such as ATAC-seq and ChIP-seq could reveal genome-wide changes in the chromatin state even in one AMF isolate exposed to several varying conditions. The results from these experiments would inform on which are the conditions that more likely trigger an epigenetic response, and then more refined experimental designs would allow to further investigate epigenetic variability, even among dikaryons, where the nuclear ratios may represent an additional source of variation.

Challenges and future perspectives in AMF research

The study of the AM symbiosis promises to revolutionize agriculture by fine-tuning plants acquisition of nutrients, and conferring resistance against pathogens or stress conditions. Evidence of such beneficial effects has been extensively documented, but to bring these findings to a practical application is an unattained milestone. The variability of AMF effects on plant growth is enormous and, at the moment, unpredictable. Efforts to understand what generates such variability have gradually narrowed down the possible sources of variation in AMF, that yet continued to display significantly different effects on plants. The fact that AMF variation is important for plant growth is very promising but at the same time, without predictability of knowing how to use this variation currently limits practical applications.

The work in this thesis attempted to unravel how genetic and epigenetic variation among clonally produced AMF lines contributed to AMF phenotypic variability, by influencing their gene expression. We showed that homokaryon clonal isolates like C2 and C5 could differ in their epigenomes; and that dikaryon SSSLs differed in their nuclear ratios which, in turn, affected their transcriptomes and germination. However, we fell short in linking AMF genetic and epigenetic variability to specific phenotypes and to the effects they have on plants.

We faced several limitations that have historically challenged the advance of AMF research, that also limited the scope of our studies. In order to build a solid bridge between genotypes and phenotypes, it is necessary to first have well defined genotypes and phenotypes. While genotyping has become easier in the recent years thanks to the advent of high-throughput DNA sequencing, phenotyping AMF is still an arduous task. Very few fungal traits can be measured precisely, and at very small scale. Forward genetics approaches, that aim to establish the genetic basis of a trait in an unbiased manner,

typically require large cohorts and well-defined phenotypes, two conditions difficult to meet in most AMF scientific studies. On the other hand, classical reverse genetics methods, that allow to link a specific gene sequence to a phenotype, are of very difficult application in AMF because of their multinucleated nature. As a result, the vast majority of the genes in AMF remain of unknown function and their metabolic pathways and gene regulatory networks are practically unexplored, with the exception of a few genes and processes involved in nutrient transport, and sugars and lipids metabolism. Fortunately, the rapid development of new sequencing technologies and sequencing strategies; the maturing of precise molecular tools for genetic engineering such as CRISPR-Cas9 editing; the establishment of effective siRNAs protocols that allow to silence specific AMF genes post-transcriptionally via genetically modified plant hosts; and the implementation and deployment of artificial intelligence algorithms in computer vision, that could allow the high-throughput phenotyping of AMF; will all hopefully help to circumvent the historical challenges. Undoubtfully, breakthroughs in AMF biology will continue to develop and are likely to advance the field in an unprecedented manner in the near future.

References

Allis CD and Jenuwein T (2016) The molecular hallmarks of epigenetic control. *Nature Reviews Genetics* 17(8): 487-500.

Angelard C, Colard A, Niculita-Hirzel H, et al. (2010) Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Curr Biol* 20(13): 1216-1221.

Angelard C, Tanner CJ, Fontanillas P, et al. (2014) Rapid genotypic change and plasticity in arbuscular mycorrhizal fungi is caused by a host shift and enhanced by segregation. *ISME J* 8(2): 284-294.

Asgari S (2014) Epigenetic modifications underlying symbiont-host interactions. *Adv Genet* 86: 253-276.

Bewick AJ, Hofmeister BT, Powers RA, et al. (2019) Diversity of cytosine methylation across the fungal tree of life. *Nature Ecology & Evolution* 3(3): 479-490.

Ceballos I, Mateus ID, Peña R, et al. (2019) Using variation in arbuscular mycorrhizal fungi to drive the productivity of the food security crop cassava. *bioRxiv*. DOI: 10.1101/830547. 830547.

Dallaire A, Manley BF, Wilkens M, et al. (2021) Transcriptional activity and epigenetic regulation of transposable elements in the symbiotic fungus Rhizophagus irregularis. *Genome Res* 31(12): 2290-2302.

Ehinger MO, Croll D, Koch AM, et al. (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(3): 853-861.

Gilbert SF, McDonald E, Boyle N, et al. (2010) Symbiosis as a source of selectable epigenetic variation: taking the heat for the big guy. *Philos Trans R Soc Lond B Biol Sci* 365(1540): 671-678.

Gladfelter AS (2006) Nuclear anarchy: asynchronous mitosis in multinucleated fungal hyphae. *Curr Opin Microbiol* 9(6): 547-552.

Gladfelter AS, Hungerbuehler AK and Philippsen P (2006) Asynchronous nuclear division cycles in multinucleated cells. *J Cell Biol* 172(3): 347-362.

Kokkoris V, Chagnon PL, Yildirir G, et al. (2021) Host identity influences nuclear dynamics in arbuscular mycorrhizal fungi. *Curr Biol* 31(7): 1531-1538 e1536.

Masclaux FG, Wyss T, Mateus-Gonzalez ID, et al. (2018) Variation in allele frequencies at the bg112 locus reveals unequal inheritance of nuclei in a dikaryotic isolate of the fungus Rhizophagus irregularis. *Mycorrhiza* 28(4): 369-377.

Mondo SJ, Dannebaum RO, Kuo RC, et al. (2017) Widespread adenine N6-methylation of active genes in fungi. *Nature Genetics* 49(6): 964-968.

Peña R, Robbins C, Corella JC, et al. (2020) Genetically Different Isolates of the Arbuscular Mycorrhizal Fungus Rhizophagus irregularis Induce Differential Responses to Stress in Cassava. *Frontiers in Plant Science* 11.

Robbins C, Cruz Corella J, Aletti C, et al. (2021) Generation of unequal nuclear genotype proportions in Rhizophagus irregularis progeny causes allelic imbalance in gene transcription. *New Phytol* 231(5): 1984-2001.

Roca MG, Kuo HC, Lichius A, et al. (2010) Nuclear dynamics, mitosis, and the cytoskeleton during the early stages of colony initiation in Neurospora crassa. *Eukaryot Cell* 9(8): 1171-1183.

Vandenkoornhuyse P, Quaiser A, Duhamel M, et al. (2015) The importance of the microbiome of the plant holobiont. *New Phytol* 206(4): 1196-1206.

UNIL | Université de Lausanne
Faculté de biologie
et de médecine
Department of Ecology and Evolution