Chromosome level genome assemblies of *Rhizophagus irregularis* strains reveal genome evolution and nuclear structure in arbuscular mycorrhizal fungi

Gokalp Yildirir

Thesis submitted to the University of Ottawa

in partial fulfillment of the requirements for the

Ph.D degree in Biology

Department of Biology

Faculty of Science

University of Ottawa

Abstract

The quality of the available genome assemblies directly dictates the quality of the genome analyses one can perform on a given species. If the aim is to study the mode of evolution and to discover the genome biology of an organism, ideally chromosome level genome assemblies are required to obtain the best results. Arbuscular mycorrhizal fungi (AMF) are one of these organisms that suffered from the lack of high-quality reference genome assemblies, which hindered in-depth analyses performed on these organisms. Despite the lack of high-quality genome assemblies, AMF symbionts have been a research focus for many decades in fungal genomics. These obligate symbionts colonize the roots of vascular plants and increase the overall root uptake of their hosts with the help of their extended hyphae. AMF have a unique cellular biology. Their cells are multinucleated with thousands of nuclei moving through their aseptate hyphae, and it was recently shown that depending on the strain, AMF nuclei either contain very similar genomic material (AMF homokaryons) or contain one of the two parental haplotypes that coexist in the same organism (AMF dikaryons, or AMF heterokaryons) in *Rhizophagus irregularis*, a model AMF species. It is hypothesized that a cryptic sexual life cycle between AMF homokaryons and dikaryons might be the driving factor behind the evolutionary success of these mutual symbionts.

My research aims to construct high quality reference genome assemblies to study the genome biology and mode of evolution of AMF strains. To achieve this, I build my thesis into three chapters. In the first chapter, I propose parasexual and sexual events that might take place in AMF dikaryons which might contribute to increased genome variability. In the second chapter, I present the first chromosomal level genome assemblies constructed for AMF homokaryons and reveal a two-compartment genome structure in *R. irregularis* nuclei. In my

third and final chapter, I construct parental haplotypes of AMF dikaryons, reveal a nuclear structure like that I identified in other AMF homokaryons, and compare the parental haplotypes to each other and to other AMF homokaryons. Overall, this study provides new reference genome assemblies for AMF homokaryons and dikaryons and reveal their interactions on a genome level. Furthermore, it provides new insights into their nuclear structure, as well as their genome biology and evolution. These discoveries help us investigate the factors that drive the evolutionary and ecological success of these plant symbionts and help us understand which possible nuclear mechanisms generate increased genetic diversity in AMF isolates.

Résumé

La qualité des assemblages génomiques disponibles dicte directement la qualité des analyses génomiques que l'on peut effectuer sur une espèce donnée. Si l'objectif est d'étudier le mode d'évolution et de découvrir la biologie du génome d'un organisme, idéalement des assemblages de génomes au niveau chromosomique sont nécessaires pour obtenir des résultats optimaux. Les champignons mycorhiziens à arbuscules (CMA) sont l'un de ces organismes qui ont eu très peu d'assemblages génomiques de référence de haute qualité, ce qui nuit aux analyses approfondies effectuées sur ces organismes. Malgré le manque d'assemblages génomiques de haute qualité, les CMA sont au centre de la recherche depuis de nombreuses décennies en génomique fongique. Ces symbiotes obligatoires colonisent les racines des plantes vasculaires et augmentent l'absorption globale des racines de leurs hôtes à l'aide de leurs hyphes étendus. Les CMA ont une biologie cellulaire unique. Leurs cellules sont multinucléées avec des milliers de noyaux se déplaçant à travers leurs hyphes aseptés, et il a été récemment démontré que selon la souche, les noyaux sont soit génétiquement très similaires (CMA homocaryons), soit ils représentent l'un des deux haplotypes parentaux qui coexistent dans le même organisme (CMA dicaryons) chez *Rhizophagus irregularis*, l'espèce de CMA modèle. On émet l'hypothèse qu'un cycle de vie sexuel cryptique entre les homocaryons et les dicaryons CMA pourrait être le facteur déterminant du succès évolutif de ces symbiotes racinaires.

Mes recherches visent à construire des assemblages de génomes de référence de haute qualité pour étudier la biologie du génome et le mode d'évolution des souches d'CMA. Pour y parvenir, je construis ma thèse en trois chapitres. Dans le premier chapitre, je propose des événements parasexuels et sexuels qui pourraient avoir lieu dans les souches dicaryotiques et qui pourraient contribuer à une variabilité accrue du génome. Dans le deuxième chapitre, je présente

les premiers assemblages génomiques au niveau chromosomique construits pour les souches homocaryotiques et révèle une structure génomique à deux compartiments dans les noyaux de *R. irregularis*. Dans mon troisième et dernier chapitre, je construis des haplotypes parentaux des souches dicaryotiques, révèle une structure nucléaire comme celle que j'ai identifiée dans les homocaryons CMA, et compare les haplotypes parentaux entre eux et à d'autres souches homocaryotiques. Dans l'ensemble, cette étude fournit de nouveaux assemblages génomiques de référence pour les souches homocaryotiques et dicaryotiques et révèle leurs interactions au niveau du génome. En outre, elle fournit de nouvelles informations sur leur structure nucléaire, ainsi que sur la biologie et l'évolution de leur génome. Ces découvertes offrent un aperçu important des facteurs qui déterminent le succès évolutif et écologique de ces symbiotes racinaires et nous aident à comprendre quels mécanismes nucléaires génèrent une diversité génétique accrue chez les CMA.

Table of Contents

Dedication

This thesis is dedicated to my wonderful wife, Bengisu Molyer Yildirir. You shaped me into a better scientist and a better man. You are the plant host to my obligate symbiont.

We will make a forest one day.

Acknowledgements

First and foremost, I would like to thank Dr. Nicolas Corradi for his continuous encouragement and excellent supervision over the last six years. Thank you for accepting me as an international volunteer student when I was just a third-year undergraduate student from Turkey and thank you for teaching me how a real scientist overcomes obstacles. Similarly, I would like to acknowledge the members of my Thesis Advisory Committee; Dr. Franck Stefani and Dr. Nicolas Rodrigue. Your insightful advice throughout the years has shaped this project and helped me become a better scientist.

A huge thank you past and present Corradi lab members, my friends, for always making me feel home and always supporting me when life gets hard. Dr. Madhu Malar C, thank you for teaching me everything I know about bioinformatics, I couldn't ask for a better mentor and a better friend. Thank you, Dr. Vasilis Kokkoris, for providing me hours of priceless discussion and advice, I will always look up to you as an excellent scientist and an exceptional cook. Dr. Jana Sperschneider, your invaluable analyses contributed to my thesis enormously, I am extremely lucky for our collaboration. A special thank you to Essam Sorwar, Matthew Villeneuve-Laroche and Calvin Cornell, I couldn't have finished this thesis without your assistance.

Lastly, I would like to thank my parents Neshan and Mahmut Yildirir, my sister Zeynep Yildirir and my wife Bengisu Molyer Yildirir. Annem ve babam, bana çalışkanlık ve dürüstlükle geçmiş bir hayatta, sınırlı maddi kaynakların asla rüyaların önüne geçemeyeceğini gösterdiniz. Size minnettarım. Zeynep, bana rüya kurmayı yeniden öğrettiğin için teşekkürler. Sevgili karım, gerçek bir bilim insanının ne kadar azimli olması gerektiğini bana defalarca gösterdiğin için çok teşekkürler.

List of Figures

***irregularis*.** In these heat maps, created to demonstrate contact frequencies throughout the genome, genome coordinates are represented on both axes. In an individual chromosome, regions that interact with each other result in an increase in contact frequencies and reveals the compartmentalized nature of the chromosome. These bright squares highlight increased contact frequency within and between chromosomes, are further analyzed to group them into euchromatin or heterochromatin compartments. Left: For chromosome l, the compartment shown in the red square belongs to compartment A, and is surrounded by two B-compartment regions shown in orange squares. Right: when the interactions of several chromosomes are analyzed, a "checkered" pattern indicates the genome arrangement of the euchromatin or heterochromatin compartments of *R. irregularis*. Telomeres are clearly visible in the Hi-C maps at the tip of each chromosome. ……40

Figure 2.3: A/B compartments have different gene (Gene/10Kb) and repeat (Repeat/10Kb) densities, and their genes contain different pfam domains. A) Upper panel: The Box-plot ranges from 0-10 and 0-9 for the A and B-compartments. Lower: The Box-plot ranges from 1-49 for both compartments. Box edges show the third quartile and first quartile in box plots, median is shown with a middle lane and the range of the data is indicated by the whiskers. Outliers are shown as dots. Asterisks indicate T-test statistical significance at $p < 0.0001$. **B) Gene numbers that carry specific Pfam domains also vary between compartments**. Circles highlight the total of number of genes carrying specific Pfam domains in located in compartment A (red circles) or B (blue circles), with the size of the circle being proportionate to the number of genes that carry that domain. Inter-strain variability in pfam domains is also evident within each compartment……..43

Figure 2.4: Gene expression and methylation analyses for the A/B compartments. **A) Genes in the A compartment show significantly higher expression levels than genes in the B-compartment in all three conditions.** Compartment A and B boxplot values range between 0-9.81 and 0-10.12 for in planta *Medicago* condition, between 0-9.09 and 0-10.53 for in planta *Allium* condition and 0-9.93 and 0-8.27 in germinated spores, respectively. **B) Genes encoding secreted proteins are up-regulated in both in planta conditions in the B-compartment, but not in the A-compartment. a)** The A-compartment shows significant upregulation of secreted proteins *in planta* Allium condition when compared to germinated spores. For A-compartment, non-secreted protein boxplot values range between -11.17 and 15.78, while secreted protein boxplot values range between -4.40 and 15.47 when in planta *Allium* expression is compared to the expression of germinated spores. **b)** The A-compartment does not show a significant secreted protein expression change when *in planta* Medicago expressions are compared to the expression of germinated spores. For A-compartment, non-secreted protein boxplot values range between -13.22 and 17.66, while secreted protein boxplot values range between -8.44 and 13.06 when in planta *Medicago* expression is compared to the expression in germinated spores. **c)** The B-compartment shows significant upregulation of secreted proteins *in planta* Allium condition when compared to germinated spores. For B-compartment, non-secreted protein boxplot values range between -11.17 and 15.08, while secreted protein boxplot values range between -5.25 and 14.46 when in planta *Allium* expression is compared to the expression in germinated spores. **d)** The B-compartment B shows significant upregulation of secreted proteins *in planta* Medicago condition when compared to germinated spores. For B-compartment, non-secreted protein boxplot values range between -13.22 and 17.66, while secreted protein boxplot values range

List of Tables

Introduction

Arbuscular mycorrhizal fungi (AMF) are mutualistic fungal species belonging to the subphylum Glomeromycotina. These fungi form symbiosis with the roots of most of the vascular plant species (Schüßler *et al.*, 2001; Corradi & Bonfante, 2012; Spatafora *et al.*, 2016; Naranjo-Ortiz & Gabaldón, 2019). AMF improve their hosts' overall root uptake of water, phosphorus, phosphate and nitrogen and transfer these nutrients to the plant host via specialized structures called arbuscules inside the root parenchyma cells. In exchange, the plant provides carbon sources such as sugars and fatty acids to the fungus (Luginbuehl *et al.*, 2017; Rich *et al.*, 2017). As such, the fungal partner is completely dependent on the host for nutrition and survival.

The Arbuscular mycorrhizal (AM) symbiosis generally boosts the growth rate of the plant host (Hoeksema *et al.*, 2010). It was also shown to protect them from various environmental stress conditions (Evelin *et al.*, 2009), including drought and plant pathogens (Bonfante & Genre, 2010). The AM symbiosis can be observed worldwide in all continents, including Antarctica (Barbosa *et al.*, 2017). AMF are also ancient, as oldest AM fossils date back to Early Devonian era, 400 million year ago; an ancient relationship that is hypothesized to be one of the driving factors leading to land colonization by plants with the help of all the benefits plants obtain from this relationship (Pirozynski & Malloch, 1975; Selosse & le Tacon, 1998). Notably, the presence of AMF can also increase the overall biodiversity and productivity of terrestrial ecosystems (van der Heijden *et al.*, 1998, 2015; Gamper *et al.*, 2010) and, as a result, today, AMF inoculum is being used in modern agriculture as a "biostimulants" to improve crop health and yield globally (Gianinazzi & Vosátka, 2011).

Cellular biology and multinucleated nature of AMF

AMF have a unique cellular structure. Specifically, their spores and mycelium constantly harbor hundreds to thousands of nuclei in a common coenocytic cytoplasm and, as opposed to other multinucleate fungi, single or bi-nucleated stages have never been observed in their spores or aseptate hyphae (Kamel *et al.*, 2017a; Kokkoris *et al.*, 2020). While other organisms, including other fungi can be multinucleated, the very large number of nuclei that coexist in a single cell sets AMF apart from other any other organism known to date (Steinkraus & Kramer, 1989; Kokkoris *et al.*, 2020).

AMF clonal reproduction and propagation begins from their globular shaped spores. In a spore germination event, numerous nuclei found within a single spore start migrating to form the germ tubes, which kickstarts the formation of the hyphal network. If a physical disturbance disrupts the hyphal structure or if the plant host cannot be encountered following germination, the nuclei that did not migrate to the germ tubes can restart the germination event (Koske, 1981; Logi *et al.*, 1998; Giovannetti *et al.*, 2000). This re-germination event can also trigger the retraction of cytoplasm and nuclei back into the spore, leaving behind a hollow hypha that is separated from the viable hyphae by a "retention septa". This strategy conserves cytoplasmic matter and nuclei, increasing the survival chance and overall fitness of the AMF spore. Following post germination and symbiosis stages, nuclear divisions increase the number of nuclei (Bécard & Pfeffer, 1993). As the new spores start to form, these nuclei then start to flow freely in and out, slowly increasing spore nuclear content (Jany & Pawlowska, 2010).

AMF genetics

The obligate symbiotic nature of AMF, combined with their multinucleate content of their cells have severely hampered our understanding of their genetics for many decades. Performing genome studies on these organisms to understand how they managed to be so ecologically and evolutionarily successful has been a research focus for the research community. For many decades, these fungi have been assumed to be "ancient asexuals" because no reproductive structures have been formally observed in these organisms (Judson & Normark, 1996). The term "ancient asexual" normally refers to taxa known to date hundreds of millions of years through analyses of fossil record but have yet to show evidence of sexual reproduction – i.e. meiosis to generate a recombined progeny from two parental strains – or sexual structures. The presence of sexual reproduction is a key aspect in evolutionary theory, as this mechanism is essential to remove deleterious mutations from a population, thus ensuring the survival of the species while simultaneously increasing individual adaptability to environmental change. In summary, without sex, populations and species should rapidly go extinct, and as a result of this dilemma, AMF have also been referred to as "asexual scandals".

Without the evolutionary advantage of sexual reproduction, how these organisms have evolved to be compatible with the majority of the land plants and how they evolved into multiple taxa to support land plants across various ecosystems and environmental stress conditions has been a great mystery. To address this, it was proposed that that AMF may have successfully evolved without sex because of an unusual genetic system whereby thousands of highly diverging/recombining nuclei co-exist in the cytoplasm. Specifically, this hypothesis (the AMF "heterokayosis" hypothesis) proposed that any mutation accumulating in some nuclei could be complemented by other co-existing nuclei (Ehinger *et al.*, 2009; Sanders & Croll, 2010; Boon *et al.*,

2015), allowing AMF to accumulate numerous mutations without the need for sexual reproduction for at least 400 million years. While this hypothesis was accepted by many for over a decade, some concerns regarding its validity also emerged. For instance, using several molecular approaches, it was found that AMF contained highly similar nuclei (Pawlowska & Taylor, 2005), and the "heterokaryosis" hypothesis was further challenged by the first whole genome sequencing studies conducted on *Rhizophagus irregularis* (a model AMF species). These first studies found little evidence of genetic diversity within the multinucleate mycelium using in-silico approaches (Tisserant *et al.*, 2013; Lin *et al.*, 2014). Many more studies challenging the notion that AMF are ancient asexuals followed on the heels of the first whole genome sequencing studies of these fungi. These studies included the discovery of a complete set of genes involved in meiosis and significant evidence for inter-strain recombination in AMF (den Bakker *et al.*, 2010; Halary *et al.*, 2011; Riley *et al.*, 2014).

Perhaps one of the strongest challenges to the notion that AMF are clonal organisms is the discovery of strains that contain the nuclear content derived from two parents (Ropars *et al.*, 2016; Corradi & Brachmann, 2017), as seen in sexual basidiomycetes and ascomycetes (Casselton, 2008). Specifically, Ropars and colleagues studied the genome structure of several strains of the model species *R. irregularis* and analyzed their single nucleotide polymorphism (SNP) frequencies, copy number of MAT-loci and ploidy levels of the strains. These analyses showed that *R. irregularis* strains can be classified into two groups: AMF homokaryons or AMF dikaryons (or AMF heterokaryons). AMF homokaryons harbor one dominant genotype in all their nuclei, and one copy of a locus that resembles the mating-type (MAT) region of sexual fungi. The MAT-locus is a region that governs sexual identity and reproduction in many fungi. In contrast to AMF homokaryons, AMF dikaryons contain multiple nuclei that are derived from

two AMF homokaryons, presumably as a result of nuclear exchange via somatic or sexual compatibility. Each nucleus also contains one of two divergent loci with similarities to the mating-type (MAT) locus of basidiomycetes (Ropars *et al.*, 2016).

These findings indicate that a cryptic sexual (or parasexual) life cycle might take place between homokaryotic and dikaryotic strains of *R. irregularis* where sexually compatible homokaryotic strains undergo plasmogamy to form dikaryotic strains, and dikaryotic strains undergo karyogamy and meiosis to form new homokaryotic strains (**Figure 0.1)**. This cryptic sexual life cycle might have created opportunities for AMF to generate genetic diversity in the absence of formal sexual reproduction, and helped them adapt to environmental changes over the past millions of years.

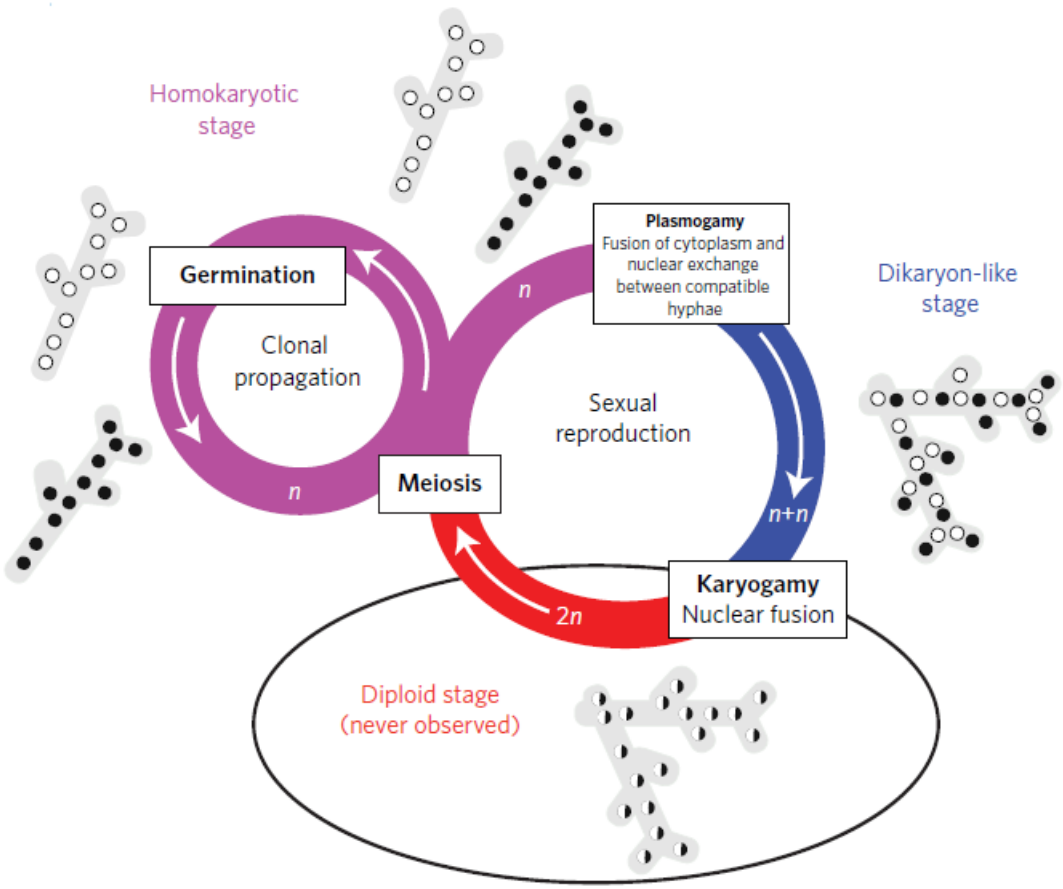

Homokaryotic stage
Germination
Clonal propagation
n
Plasmogamy
Fusion of cytoplasm and nuclear exchange between compatible hyphae
Dikaryon-like stage
Sexual reproduction
n+n
Meiosis
n
2n
Karyogamy
Nuclear fusion
Diploid stage
(never observed)

Figure 0.1: Model representing the putative sexual life cycle of *Rhizophagus irregularis*, a model AMF species. Two sexually compatible homokaryotic strains undergo plasmogamy to produce a dikaryotic strain. Following karyogamy and meiosis, the dikaryotic strain gives rise to two new homokaryotic strains. Figure taken from Ropars et al, 2016.

AMF genome analyses

Genome analyses can provide essential insights onto the biology of any organisms, particularly hard to culture microorganisms, by uncovering which biochemical pathways are missing or have expanded in their genomes. In the case of AMF genomes, genome analyses revealed that these fungi have lost essential genes involved in fatty acid and carbohydrate metabolism, highlighting why they need those compounds from their hosts. Furthermore, phylogenomics showed that these losses occurred before the emergence of the Glomeromycotina subphylum (Sun *et al.*, 2019; Morin *et al.*, 2019; Malar C *et al.*, 2021). Despite these key gene losses, AMF genomes are still among the largest fungal genomes, ranging from 44Mb (*Paraglomus occultum*, (Malar C *et al.*, 2022)) to 568Mb (*Gigaspora rosea*, (Morin *et al.*, 2019)) within the range of fungi with massive genomes such as *Tuber melanosporum* and powdery mildews (Martin *et al.*, 2010; Spanu *et al.*, 2010; Tisserant *et al.*, 2013). One factor contributing to this genome expansion is the high repeat content of these genomes. Recent genome studies discovered that at least 25% of *R. irregularis* genome is composed of repeats, and these repeats are mainly composed of transposable elements (Chen *et al.*, 2018b). Notably, most of these transposable elements are transcribed, suggesting that they continue contributing to the genome size expansion in this subphylum. However, exceptions to the transposable element abundance in AMF genomes exist, and indeed one AMF species containing relatively few repeats and smaller genome sizes (40Mb) is also known (*Paraglomus occultum*, (Malar C *et al.*, 2022)).

Until now, about 300 AMF species have been described based on the spore morphologies, but ribosomal DNA sequences show that up to 1600 operational taxonomic units may exist in nature (Öpik & Davison, 2016). As the morphological diversity is restricted, it is intriguing how AMF can have such a broad host range that they can interact with 200 000

different plant species (Rodríguez-Echeverría *et al.*, 2017). Studies targeting *R. irregularis* discovered significant phenotypical variability in hyphal length and spore count between different strains that were collected from the same field (Koch *et al.*, 2006; Croll *et al.*, 2008). These strains also differ from each other substantially on a genetic level, sharing as little as 50% of their gene repertoire (Chen *et al.*, 2018b). Notably, many pathways presumably involved in host interactions - i.e. dispensable or accessory genes – differ significantly in presence and in copy numbers among strains of the same species. This huge variation in gene content indicate that each strain of an AMF species has unique molecular functions, some of which could be involved in plant host interactions.

Another intriguing aspect of the AMF genomes is how the abundance of transposable elements might trigger events that result in increased genome diversity between different strains. The abundance of transposable elements is a known indicator of large-scale genome organization events that took place (Seidl & Thomma, 2014; Faino *et al.*, 2016) in fungal pathogens, and it is intriguing to speculate that similar mechanisms can also occur in AMF. Indeed, the degree of variability between AMF strains is also in repeat content, with some transposable elements being abundant in some strains but not in others. The transposable element variability, combined with the gene content variability might indicate a "two-speed genome evolution" model in this fungal group as seen in other fungal pathogens, where genomic regions enriched with transposable elements evolve more rapidly (Faino *et al.*, 2016).

The term "pangenome" originally coined to describe the gene diversity in prokaryotes (Medini *et al.*, 2005; Marroni *et al.*, 2014; Vernikos *et al.*, 2015), and this term has recently been introduced to describe intra-specific genome variability in eukaryotes, including fungi, and more recently AMF (Chen *et al.*, 2018b; Mathieu *et al.*, 2018). Assuming that analyzed strains indeed

belong to the same species, the pangenome of *R. irregularis* is one of the largest known for eukaryotes, since at the extreme only half of the gene content is shared between conspecific strains. In fact, the number of genes contributing to the *R. irregularis* pangenome altogether exceeds 150,000 after analyzing only six strains (Chen *et al.*, 2018b), which suggests that this model AMF carries a massive gene diversity of accessory genes. These accessory genes are hypothesized to code for molecular mechanisms that provide ecological and evolutionary advantage to the strains, helping them adapt to changes in the environment more efficiently, or boost a strain's symbiosis range with the plant hosts.

The necessary next steps to fully understand AMF genetics and genomes

To date, most available AMF genome assemblies are fragmented into thousands of contigs. This fragmentation prevents further discovery of synteny blocks and structural variants between different strains that are needed to build strong hypothesis about the genome evolution of these organisms. Indeed, high fragmentation rate in a genome assembly can introduce many artefacts, including collapsed regions, chimeric assemblies, and artificial gene duplications (Denton *et al.*, 2014; Montoliu-Nerin *et al.*, 2020) which prevent proper analysis of genome rearrangements, repeat diversity and repeat plasticity across the strains. In other fungal pathogens, high repeat content is shown to boost intraspecific genetic variability, and recently, it was proposed that similar mechanisms that are triggered by high repeat content might cause increased diversity in AMF genomes (Chen *et al.*, 2018b; Mathieu *et al.*, 2018). However, to further test the link between high repeat content and high intraspecific genome diversity, ideally chromosome level, reference genome assemblies need to be analyzed.

High fragmentation rate of the genome assemblies also prevents a deeper analysis of AMF genome biology. In other eukaryotic organisms, chromosomes are folded into two groups, euchromatin and heterochromatin. Each compartment is regulated; euchromatin compartment has a higher gene expression rate due to its open structure and heterochromatin compartment has a lower gene expression as a result of its compacted structure (Pombo & Dillon, 2015; Dekker & Heard, 2015; Szabo *et al.*, 2019; Kempfer & Pombo, 2020). This genome compartmentalization is observed in a few other fungal species (Galazka *et al.*, 2016; Kim *et al.*, 2017; Winter *et al.*, 2018) but the epigenetic role of chromosome folding on AMF genome biology is yet to be analyzed. The lack of reliable genome assemblies also hinders the genome studies targeting the dikaryotic strains. The assemblies of AMF dikaryons harbor all the problems that studies of AMF homokaryons contain, but on top of that the two parental haplotypes coexisting in these strains cannot be separated. As such, all known assemblies from AMF dikaryons contain significantly collapsed parental haplotypes, hampering analyses of genetic diversity, as well as genome and epigenetic content among the parental haplotypes.

In true dikaryons (i.e. two nuclei per cell), the coexisting haplotypes can interact in multiple ways, including competition leading in some cases to the elimination of one genotype over time (Ryan & Lederberg, 1946; Rayner, 1991) or cooperation (exchanging genetic information between coexisting nuclei through somatic recombination events) (Pontecorvo *et al.*, 1953; Clark & Anderson, 2004). In other fungal dikaryons, two parental genotypes can express distinct gene sets along different life-stages of the life cycle, working together to support the fungus throughout its life.

Unfortunately, for AMF dikaryons, since the two parental haplotypes cannot be separated fully and assembled properly, the interactions between two haplotypes cannot be analyzed or the

origin of the two haplotypes cannot be determined to reveal how the dikaryotic state contribute to the mode of evolution of AMF. Similarly, because the genes that belong to each parental haplotype cannot be separated, their respective contribution to the overall transcriptome and genome biology of AMF dikaryons cannot be thoroughly investigated.

Research justification and goals

In this PhD thesis, I aim to increase our overall knowledge on AMF genomics and genome biology. My goal is to obtain a fine-scale view of the genome content, epigenetics structure, and mode of evolution of model AMF strains.

My first chapter is published as a review introducing new models that explain how genome diversity is generated in AMF dikaryons. These models build on top of the Ropars et al cryptic sexual life cycle model, but they require empirical testing using high quality genome assemblies, and I will detail how these can be obtained in the next chapters. For my second chapter, I construct chromosomal level genome assemblies for five homokaryotic *R. irregularis* strains to finally reveal their genome content in full scale. These analyses also provide a primary view of AMF genome biology through the discovery of two chromosomal compartments in this species, each playing a distinct role in the biology and evolution of this fungus. In my third and final chapter, I show that a phased, complete, and chromosome-level view of all dikaryotic strains can be obtained using appropriate methodologies. Analyzing the phased, or separated, parental haplotypes reveal insight into their origin and allow us to discover their relative genetic and epigenetic contribution to the mycorrhizal symbiosis. These analyses also show how these

haplotypes interact with each other genetically, and how these changes affect each AMF strain in different ways.

Altogether, this work pushes the boundaries in our understanding of AMF genomes and nuclear organization. The new reference genome assemblies produced in this study will be available to be used by other researchers to further the field of AMF genomics.

Chapter 1: Parasexual and Sexual Reproduction in Arbuscular Mycorrhizal Fungi: Room for Both

Gökalp Yildirir[1,2], Mathu Malar C[1,2], Vasilis Kokkoris[1], and Nicolas Corradi[1,*]

[1]Department of Biology, University of Ottawa, Ottawa, ON, Canada

[2]These authors contributed equally to this work.

* Corresponding Author

This work has been published in Trends in Microbiology, July 2020: Yildirir, G., Malar C, M., Kokkoris, V., & Corradi, N. (2020). Parasexual and Sexual Reproduction in Arbuscular Mycorrhizal Fungi: Room for Both. *Trends in microbiology*, *28*(7), 517–519. https://doi.org/10.1016/j.tim.2020.03.013

Summary

Arbuscular mycorrhizal fungi (AMF) harbor thousands of nuclei in a large syncytium at all times. Although mating processes have not been observed in AMF, their cells and genomes show many signatures of sexual reproduction. Here, we describe how some of these signatures could also arise from parasexual processes in these widespread plant symbionts. As such, parasexual and sexual evolution could both be at play in generating nuclear diversity in AMF.

Introduction

The notion that arbuscular mycorrhizal fungi (AMF) are ancient asexual organisms has been challenged by recent genomic analyses (Ropars *et al.*, 2016). Specifically, genomic regions and nuclear organizations linked with sexual reproduction in fungi are found in these organisms. These include the presence of meiosis-specific genes, putative mating-type loci, and homokaryotic-dikaryotic life stages. Here, we argue that various parasexual mechanisms can also generate significant genetic diversity in AMF.

Meiotic Genes and Parasexuality

Evidence that individual AMF strains can recombine genetic material with one another has been found using microsatellites and single-gene sequence data and through whole-genome analyses (e.g., (Vandenkoornhuyse *et al.*, 2001; Croll & Sanders, 2009; Chen *et al.*, 2018b)). This demonstrated that AMF have found ways to diversify their genomes in the absence of observable sex; yet, it is still unclear when recombination emerges by sexual or parasexual means.

One of the largest drivers of recombination in eukaryotic organisms is meiosis. This process defines sexual reproduction in eukaryotes, as it shuffles the genetic material between homologous chromosomes following nuclear fusion (karyogamy) (Heitman *et al.*, 2007). AMF harbor a full set of meiosis-specific genes in their genomes (Halary *et al.*, 2011), suggesting these fungi can theoretically undergo meiosis. To this day, however, key meiosis events such as karyogamy have not been observed, and genetically distinct strains created by meiotic divisions have never been identified. This raises the question: are meiosis-specific genes always involved in sexual reproduction in AMF?

For some time, the ascomycete fungus *Candida albicans* presented a similar dilemma. Like AMF, this species carries all proteins necessary for meiosis; yet, this cellular process has never been observed in this species. New findings are now revealing why; in *C. albicans*, key meiosis genes (e.g., Spo11,Rec8) are actually involved in previously unsuspected parasexual events (Anderson *et al.*, 2019). This process, named 'parameiosis', produces high recombination rates without the need for sex. Although parameiosis is unlikely to occur in AMF (i.e., it requires karyogamy between diploid nuclei, and all nuclei analyzed to date in AMF are haploid), its discovery raises the intriguing possibility that AMF also use their meiotic machinery to recombine nuclei through parasexual means. In the arbuscular mycorrhiza dikaryons, such processes could generate some of the rare recombination events reported by many (Vandenkoornhuyse *et al.*, 2001; Croll & Sanders, 2009; Chen *et al.*, 2018b), similarly to what has been observed in distant dikaryotic fungi (Anderson & Kohn, 2014). In summary, while meiotic genes are conserved in AMF, their function remains unclear at this time. It is thus possible that, in addition to meiosis, meiotic proteins also control parasexual/noncanonical genetic pathways to generate diversity and address environmental challenges.

How Could Meiosis Function in Dikaryotic AMF?

Recent work showed that AMF strains are either homokaryotic, where all nuclei contain one type of genome, or dikaryotic, where nuclei originating from two parental strains coexist in the cytoplasm (Ropars *et al.*, 2016). A model explaining these findings based on current knowledge of sexual reproduction in other fungi was proposed. Briefly, the model hypothesizes that AM dikaryons are progenies of sexually compatible homokaryotic strains, which could undergo meiosis at some point to produce a recombined homokaryotic spore progeny. This hypothesis is based on genomic data and knowledge of fungal sexual compatibility, but it has not yet been experimentally validated. If sexual compatibility exists in AMF, it is unclear how meiosis can be regulated among the thousands of nuclei that flow simultaneously in their cells. In such a unique genetic system, meiosis either occurs independently at different locations within the mycelium or is somehow coordinated among all coexisting nuclei. Alternatively, meiosis can take place only in cases where plasmogamy produces spores with only two sexually compatible nuclei (i.e., like in conventional dikaryotic cells), but to date the presence of one or two nuclei in single spores has never been observed in AMF. Either way, the meiotic machinery of AMF must have evolved new ways to cope with a complex genetic system, and experimental procedures should now be developed to address this question.

Opportunities for Somatic Hybridization/Recombination in Dikaryotic Strains

Assuming that AMF never produce binucleated spores/cells, can their dikaryotic stages offer opportunities for recombination without sex? Here, we argue that parasexual mechanisms such as somatic recombination and/or transposable element (TE)-induced genetic transfers could achieve this by taking advantage of the constant multinucleate state of AM dikaryons (Figure 1.1).

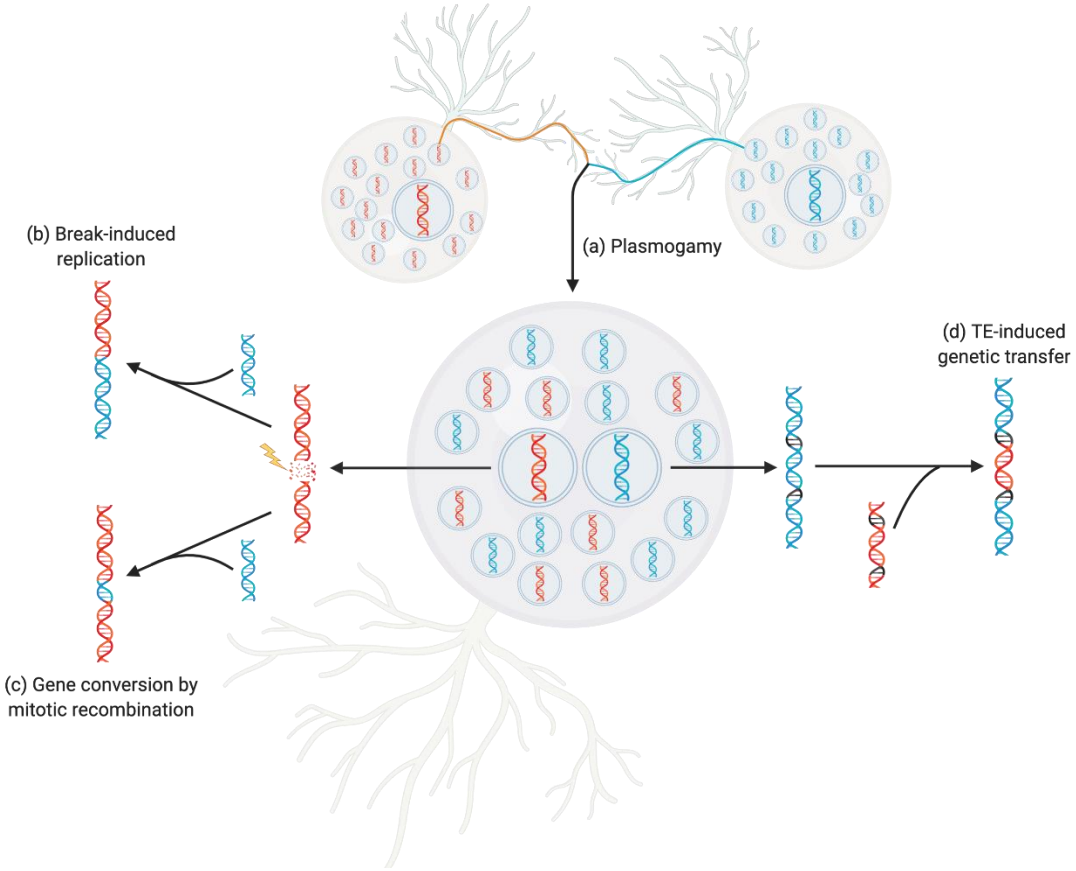

(b) Break-induced replication
(a) Plasmogamy
(d) TE-induced genetic transfer
(c) Gene conversion by mitotic recombination

Figure 1.1 Potential Parasexual Mechanisms Driving Genetic Diversity in Arbuscular Mycorrhizal Fungi. Parental genomes found in arbuscular mycorrhiza (AM) dikaryons are shown as red/orange or blue DNA strands, and transposable element (TE) repeat regions are shown in black. **(A) Plasmogamy between two compatible homokaryotic strains and putative parasexual events.** A nuclear exchange (plasmogamy) between two compatible homokaryotic strains gives rise to a dikaryotic strain that contains both parental haploid nuclei in similar frequencies. We propose that very close spatial and genetic proximity of nuclei in AM dikaryons can fuel a number of parasexual events. **(B) Break-induced replication (BIR) and internuclear genetic exchange.** A double-strand break in DNA structure can trigger BIR. The template is used to reconstruct the modified chromosome from the double-strand break region to the end. In AM dikaryons, BIR may occur after a karyogamy event between nuclei of different genotypes or between two nuclei that are replicating through mitosis side by side. In both cases, this repair mechanism would use a homologous chromosome from the opposite parental genome as a template, producing a recombined genotype. **(C) Gene conversion by mitotic recombination.** A double-strand DNA break can also trigger gene conversion by mitotic recombination between the homologous sequences. To occur in AM dikaryons, the homologous chromosomes of each coexisting nucleotype must have undergone either karyogamy or replication occurring nearby. In a gene conversion event, only part of the DNA is recombined between homologous genomic regions. Note that an AMF meiosis-specific gene (e.g., Spo11 and Rec8) could be involved in these repair processes. **(D) TE- mediated gene exchange.** TEs can self-replicate and insert themselves into both parental genomes. A gene region that is unique to one parental genome can thus be transferred to the other parental strain, migrating along with the flanking TE repeats. (Figure created with BioRender.com.)

As opposed to meiosis, in a parasexual recombination cycle, nuclear fusion (i.e., diploidization) is not synchronized. In this scenario, nuclei of any mating type could theoretically undergo somatic recombination at any given time, as seen in fungal dikaryons (Anderson & Kohn, 2014). Indeed, somatic re- combination is frequently observed in dikaryotic basidiomycetes following hyphal fusion, after which crossing over between parental nucleotypes can occur with each round of mitotic division (Xu *et al.*, 1996; Anderson & Kohn, 2014). Similar mechanisms explain in part the rare internuclear recombination patterns recently reported for individual nuclei sequenced from AM dikaryons (Chen *et al.*, 2018a).

TEs are selfish DNA sequences that are capable of changing their locations within and across coexisting genomes. TEs are abundant in AMF genomes (Chen *et al.*, 2018b), and the genetic proximity of thousands of nuclei in AMF cells offers obvious opportunities for these selfish elements to transfer across nuclei to generate recombination-like genetic patterns. In support of this, it was recently shown that TEs can fuel the genetic transfer of protein-coding genes across species boundaries in pathogenic fungi (McDonald *et al.*, 2019). As such, it is easy to imagine similar processes occurring (rarely) among the thousands of nuclei coexisting in the same AMF cytoplasm. TEs could also be a substrate to trigger the somatic recombination events mentioned earlier due to their elevated sequence homology. To support the hypothesis that recombination in AMF can also be TE-driven, future work should investigate if some/most of the interstrain and internuclear recombination found in AMF is located in TE-rich regions of available genomes and within sequences from individual nuclei.

Finally, it was recently shown that dikaryotic fungal lineages can also emerge through rare somatic hybridization events involving genetically divergent strains (Li *et al.*, 2019). Specifically, it was found that a newly emerged strain (Ug99) of the wheat stem rust pathogen

Puccinia graminis f. 26 sp. *tritici* carries in its dikaryotic cells, nuclei originating from two geographically distant dikaryotic strains. Both nuclei show no sign of recombination or reassortment (Li *et al.*, 2019), an indication that this dikaryon did not recombine through sex.

It is intriguing to speculate that, in cases where plasmogamy is not driven by the mating type, natural AMF populations undergo similar somatic hybridizations, providing them with other opportunities to boost nuclear diversity in the absence of a compatible partner.

Sex versus Parasex in AMF: Need for Empirical Testing

Genome and single-nucleus data carry evidence that AMF can generate genetic diversity through sexual reproduction. Still, these findings are not incompatible with the existence of parasexual processes in these organisms, so it is likely that both processes have played (and continue to play) a role in generating diversity in AMF communities. Models that aim to explain AMF genetics and functional diversity are now available (Ropars *et al.*, 2016), and these can finally be empirically tested to determine the frequency of parasexual versus meiotic processes in these widespread symbionts.

Crossing experiments involving putative compatible homokaryotic AMF strains could be used to address this. In the presence of sexual reproduction, plasmogamy should produce new spore progenies that carry fully recombined genomes originating from crossed parental homokaryons. In parallel, long-read technologies could be used to test the hypothesis that somatic recombination can occur in these plant symbionts. In this case, pairwise comparisons of complete individual nuclei of known dikaryotic strains could be generated to validate recent findings (based on short reads) of rare and likely parasexual internuclear recombination. This

approach will also reveal the frequency of internuclear recombination in AM dikaryons and its effect on the biology and genome evolution of individual nuclei. Long-read sequence data from individual nuclei would also reveal if recombination breaks were found in genomic regions surrounded by TEs, thus helping determine the existence of TE-driven internuclear transfers in AMF (McDonald *et al.*, 2019).

Chapter 2: Long reads and Hi-C sequencing illuminate the two compartment genome of the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*

Gokalp Yildirir [1*], Jana Sperschneider [2*], Mathu Malar C [1], Eric CH Chen [3], Wataru Iwasaki [3], Calvin Cornell [1] and Nicolas Corradi [1#]

[1] Department of Biology, University of Ottawa, ON, Ottawa, K1N 6N5, Canada

[2] Biological Data Science Institute, The Australian National University, Canberra, ACT 260, Australia

[3] Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, 1113-0033, Japan

* contributed equally

[#] Corresponding Author

This work has been published in New Phytologist, November 2021: Yildirir, G., Sperschneider, J., Malar C, M., Chen, E. C., Iwasaki, W., Cornell, C., & Corradi, N. (2022). Long reads and Hi-C sequencing illuminate the two-compartment genome of the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytologist*, *233*(3), 1097-1107.

Abstract

Chromosome folding links genome structure with gene function by generating distinct nuclear compartments and topologically associating domains (TADs). In mammals, these undergo preferential interactions and regulate gene expression. However, their role in fungal genome biology is unclear. Here, we combine Nanopore (ONT) sequencing with chromatin conformation capture sequencing (Hi-C) to reveal chromosome and epigenetic diversity in a group of obligate plant symbionts; the arbuscular mycorrhizal fungi (AMF). We find that five phylogenetically distinct strains of the model AMF *Rhizophagus irregularis* carry 33 chromosomes with substantial within species variability in size, as well as in gene and repeat content. Strain-specific Hi-C contact maps all reveal a 'checkerboard' pattern that underline two dominant euchromatin (A) and heterochromatin (B) compartments. Each compartment differs in the level of gene transcription, regulation of candidate effectors and methylation frequencies. The A-compartment is more gene-dense and contains most core genes, while the B-compartment is more repeat-rich and has higher rates of chromosomal rearrangement. While the B-compartment is transcriptionally repressed, it has significantly more secreted proteins and *in planta* up-regulated candidate effectors, suggesting a possible host-induced change in chromosome conformation. Overall, this study provides a fine-scale view into the genome biology and evolution of prominent plant symbionts, and opens avenues to study the epigenetic mechanisms that modify chromosome folding during host-microbe interactions.

Introduction

Arbuscular mycorrhizal fungi (AMF) are obligate plant mutualistic organisms of the fungal subphylum Glomeromycotina (Spatafora *et al.*, 2016) that evolved a symbiotic relationship

with the roots of most vascular plants species (Corradi & Bonfante, 2012; Martin *et al.*, 2017; Lutzoni *et al.*, 2018). During the AM symbiosis, these fungi enhance their host nutrient and water uptake in exchange of sugars and lipids, and provide an increased resistance against pathogens (Bonfante & Anca, 2009). The first fossil records of the AM symbiosis date several hundred million years, and today this symbiotic relationship between AMF and their plant hosts is widespread globally, highlighting AMF's capacity to persist in various environmental conditions (Delaux & Schornack, 2021). AMF spores and hyphae harbor up to thousands of coexisting nuclei at all times (Kokkoris *et al.*, 2020) and their genetics are defined by the presence of two life stages – i.e. the homokaryotic stage (AMF homokaryons), where co-existing nuclei are genetically largely uniform, and the heterokaryotic stage (AMF dikaryons), where nuclei originating from two parental strains coexist in the mycelium (Ropars *et al.*, 2016; Corradi & Brachmann, 2017; Chen *et al.*, 2018b; Kokkoris *et al.*, 2021).

Genome analyses revealed that species in Glomeromycotina carry gene losses in cellular pathways that inhibit the metabolism of fatty acids, sugars and plant cell wall degradation enzymes (Tisserant *et al.*, 2013; Lin *et al.*, 2014; Chen *et al.*, 2018b; Morin *et al.*, 2019; Malar C *et al.*, 2021) and contain an abundance of transposable elements (TE). Closely related AMF strains also vary dramatically in their gene content, suggesting that AMF have pangenomes (Chen *et al.*, 2018b; Mathieu *et al.*, 2018). The pangenome concept highlights the distinction between 'core' conserved genes hypothesized to be essential for daily activities (e.g. cytoskeleton, protein translation, etc.), and 'dispensable' non-conserved genes thought to help individual strains adapt to changing environments and hosts (e.g. transduction signalling, effectors, etc.) (Li *et al.*, 2018; McCarthy & Fitzpatrick, 2019; Badet *et al.*, 2020).

Identifying the mechanisms from which pangenomes emerge is essential for understanding the mode of evolution of these widespread plant symbionts. In fungal plant pathogens, high repeat content generates within-species genome variability, and it was proposed that similar mechanisms drive genetic variability in AMF (Chen *et al.*, 2018b; Mathieu *et al.*, 2018). However, this hypothesis is currently untestable for AMF because most available genome assemblies are highly fragmented – i.e. composed in many cases of dozens of thousands of contigs (Chen *et al.*, 2018b; Mathieu *et al.*, 2018). This high fragmentation rate can generate mosaic sequences of multiple haplotypes (collapsed regions), chimeric assemblies and artificial gene duplications (Denton *et al.*, 2014; Montoliu-Nerin *et al.*, 2020), preventing fine-scale and reliable analyses of genome rearrangements, repeat diversity, and repeat plasticity within and across strains.

The highly fragmented state of AMF genomes also hampered the identification of mechanisms that define relationships between genome structure and function. Within nuclei, chromosomes separate into two groups: active open chromatin A-compartments (euchromatin) and less active chromatin B-compartments (heterochromatin) and each compartment can be differently regulated (Pombo & Dillon, 2015; Dekker & Heard, 2015; Szabo *et al.*, 2019; Kempfer & Pombo, 2020). This compartment-driven gene regulation highlights the key epigenetic role of chromosome folding for genome biology, but knowledge of this process in fungi was thus far restricted to a few model genera (e.g. *Sacharomyces*, *Neurospora*) and the filamentous fungus *Epichloë festucae* (Galazka *et al.*, 2016; Kim *et al.*, 2017; Winter *et al.*, 2018).

In the plant symbiont *E. festucae,* chromosome folding and repetitive elements together divide the genome into distinct regions with very similar gene expression suggesting that both genome characteristics co-regulate symbiotic associations in this species. However, whether

similar mechanisms drive gene expression and symbiotic associations in other fungal symbionts, including AMF, is unknown. Here, we aimed to obtain a complete understanding of the AMF genome biology and within species genome dynamics by acquiring long-read and strain-specific high-throughput chromatin conformation capture (Hi-C) sequencing data (Schmitt *et al.*, 2016) from five phylogenetically distinct homokaryotic strains of the species *R. irregularis*.

Materials and Methods

Culturing, DNA Extraction and ONT/Hi-C Sequencing

R. irregularis strains A1 (DAOM664342), B3 (DAOM664345) and C2 (DAOM664346) were originally isolated from Switzerland (Tänikon), while the strain DAOM197198 was originally isolated from Pont Rouge, Canada (Koch *et al.*, 2004; Stockinger *et al.*, 2009) and the strain 4401 (DAOM240446) was isolated from *Ammophila breviligulata* in "La Martinique Îles-de-la-Madeleine" (Canada).

All five strains were cultured *in vitro,* using *Daucus carota* root organ cultures (ROCs) as the host, as previously described (Corradi *et al.*, 2004a). The strains propagated in two-compartment ROCs, allowing us to produce mycelium and spores without obvious contaminants. For Oxford Nanopore (ONT) sequencing, high quality, high molecular weight DNA was extracted for the strains A1, C2, B3 and 4401 using a protocol proposed by Schwessinger and McDonald (Schwessinger, 2016). Following extraction, high quality DNA samples were processed using Nanopore Ligation Sequencing Kit SQK-LSK109 to prepare the sequencing libraries, which were sequenced using the MinION R9.4.1 flowcells to produce an average of 8.45 million reads with an N50 > 4Kb, reaching to a genome coverage of 160x.

Chromosome assembly and annotation

ONT reads were basecalled with guppy (version 4.5.2) using the "high accuracy" configuration (dna_r9.4.1_450bps_hac.cfg), and were used to generate assemblies using Canu (version 2.0) (Koren *et al.*, 2017). These assemblies were then further polished using Racon (v 1.4.10) (Vaser *et al.*, 2017) using the default parameters and the consensus sequences were created using Medaka (version 1.0.3). The polishing step was concluded with two rounds of Pilon (version 1.23) using Illumina reads and default parameters (Walker *et al.*, 2014). Illumina reads were then mapped on the polished contigs using bwa mem (version 0.7.17) (Li, 2013) and the read coverage of each contig was calculated using bedtools (version 2.26.0) (Quinlan & Hall, 2010). Contigs having abnormally low (<5x) and high (>200x) coverages were filtered out using purge_haplotigs (Roach *et al.*, 2018), and the remaining contigs were used for further scaffolding. Illumina reads used for polishing from the strains A1, B3 and C2 were obtained from Ropars *et al*. (Ropars *et al.*, 2016), DAOM197198 reads were obtained from Chen et al. 2018(Chen *et al.*, 2018b), 4401 Illumina reads were used in this study for the first time.

Approximately 200 mg of fresh mycelium from each strain was crosslinked and shipped to Phase Genomics (Seattle, WA) to obtain strain-specific Hi-C data. Hi-C data was produced using the Proximo Hi-C Kit (Microbe) at the Phase Genomics with DPNII restriction enzymes. Before generating high-throughput Hi-C Illumina data, the quality of these libraries was assessed by mapping a low coverage paired-end data onto available *R. irregularis* assemblies. A library is deemed to be of high quality if the fraction of high quality paired-end reads mapping respectively > 10KB apart within and across publicly available contigs exceeds 1.5% and 5.0%. In all cases, the values obtained for our samples far exceeded the QC limits set by Phase Genomics (Seattle, WA) – e.g. average 11.75% and 23%.

The subsequent Illumina library preparation and sequencing yielded paired-end Hi-C reads of 150 bp at 75x-150x coverage. Scaffolding was carried out with contigs that were obtained using Canu and strain-specific Hi-C data. For scaffolding, the Hi-C reads were first mapped to each contig using BWA-MEM 0.7.17 (Li, 2013). Alignments were then processed with the Arima Genomics pipeline (Arima Genomics, 2019). Scaffolding was performed using SALSA 2.2 (Ghurye *et al.*, 2017, 2019) and subsequent manual curation was guided by Hi-C contact maps. For strain C2, SALSA scaffolding resulted in 202 scaffolds from 325 contigs and 15 contigs were broken based on Hi-C contact map information. For strain A1, SALSA resulted in 107 scaffolds from 276 contigs and 14 contigs were broken based on Hi-C contact map information. For strain 4401, SALSA resulted in 108 scaffolds from 306 contigs and 13 contigs were broken based on Hi-C contact map information. For strain B3, SALSA resulted in 138 scaffolds from 314 contigs and 13 contigs were broken based on Hi-C contact map information. In the strain DAOM197198 SALSA resulted in 153 scaffolds from 210 contigs, and one contig was broken based on Hi-C contact map information.

Gene annotation was achieved by using Funannotate (version 1.7.4) (https://zenodo.org/record/2604804#.YPm336iSnIU) on repeat-masked genome assemblies. Genome annotations were performed using RNA-seq reads and protein models for the strains DAOM197198, A1, B3 and C2 available from the Joint Genome Institute (JGI; (Chen *et al.*, 2018b)). For the strain 4401, genome annotation was carried by combining EST, RNA-seq and models from the strains DAOM197198, C2 and B3. The quality of annotations was evaluated using the Benchmarking Universal Single-Copy Orthologs (version 4.0, database Obd_10) (Simão *et al.*, 2015).

Conserved protein domains were predicted using Pfam v.27 (El-Gebali *et al.*, 2019). SignalP 4.1 (-t euk -u 0.34 -U 0.34)(Petersen *et al.*, 2011) and TMHMM 2.0(Krogh *et al.*, 2001) were used to predict secreted proteins. A protein was called secreted if it was predicted to have a signal peptide and but no transmembrane domains. Effector candidates were predicted with EffectorP 3.0(Sperschneider & Dodds, 2021). De novo repeats were predicted with RepeatModeler 2.0.0 and the option -LTRStruct (Flynn *et al.*, 2020). These were merged with the RepeatMasker repeat library and RepeatMasker 4.1.0 was run with this combined repeat database (Smit *et al.*, 2013). Transposable element locations were extracted from the Repeatmasker output file using an R script and simple repeats, unknown and low complexity repeats, satellites, tRNA, snRNA and rRNAs were filtered out from the output file. PfamScan (Hancock & Bishop, 2004; Li *et al.*, 2015) was used with default parameters to identify the protein domain annotations in all strains. To create the Pfam heatmaps, the domain numbers were counted for each *R. irregularis* strain or A/B compartment, and a t-test conversion was made to highlight the protein domain abundances for each category.

We used mash dist (version 2.2.0) (Ondov *et al.*, 2016) for *k-mer* distance calculations and dnadiff from the MUMmer (Kurtz *et al.*, 2004; Marçais *et al.*, 2018) suite for structural variation calculations between strains and compartments. Orthology analyses were made by FastOrtho (Wattam *et al.*, 2014), using following parameters: 50% identity and 50% coverage using protein sequences of all five assemblies. All karyoplots were produced using KaryoploteR (Gel & Serra, 2017). All genome data and reads newly obtained are available in Genbank under the BioProject PRJNA748024.

Identification of chromosome compartments and topologically associated domains

We called A/B compartments and TADs with hicexplorer 3.6 (Ramírez *et al.*, 2018) using the commands hicPCA and hicFindTADs, respectively. Inter- and intra-chromosomal Hi-C contact maps were produced with HiC-Pro 2.11.1 (Servant *et al.*, 2015) (MAPQ=10) and hicexplorer 3.6 (Wolff *et al.*, 2018, 2020; Winter *et al.*, 2018; Ramírez *et al.*, 2018) and these Hi-C contact maps were manually inspected to assign chromosomal regions to A/B compartments. Specifically, the regions along the each chromosomes carrying the same PCA1 values – i.e. positive or negative – were assigned to the same compartment. Following this assignment, compartments were manually inspected to investigate their proximity with other chromosomes. Those in close physical proximity with the smallest chromosomes were assigned to compartment A, while those carrying the counterpart PCA1 signal were assigned to compartment B. This process was manually repeated for all 33 chromosomes individually to separate the chromosomal regions into separate A/B compartments. Bedtools (Quinlan & Hall, 2010) was used to assess overlap between genomic features such as genes/repeats and compartments.

Methylation and gene expression analyses

Megalodon (version 2.2.9) was used for methylation calling using the high accuracy parameters (dna_r9.4.1_450bps_modbases_dam-dcm-cpg_hac.cfg) configuration file. After methylation calling was completed for all five assemblies, CG positions and their methylation frequencies were extracted from the 5mC bedfile output. For RNA-seq analyses, Salmon v.1.3.0 (Patro *et al.*, 2017) was used to align clean RNA-seq reads to the transcripts and to estimate transcript abundances in each sample (salmon index –keepDuplicates and salmon quant –validateMappings). We used tximport and DESeq2 (Love *et al.*, 2014) to assess gene differential

expression (padj < 0.1). The DAOM197198 RNA-seq datasets used in this study were obtained from germinating spores (SRR1979300-SRR1979302), as well as intra-radical material isolated from *Medicago trunculata* (SRR5644319-SRR5644324) and *Allium schoenopasum* (SRR5644331- SRR5644333).

Results

Assembly and annotation of Rhizophagus irregularis chromosomes

Basecalled and polished ONT data from four homokaryotic strains (A1, B3, C2 and 4401) were assembled using Canu (Koren *et al.*, 2017), and then scaffolded using strain-specific Hi-C data. The same scaffolding procedures were performed on a previously published PacBio assembly of the model AMF strain DAOM197198 (Maeda et al., 2018a)). This approach resulted in an average of 137 scaffolds with an N50 score of 5 Mb, which were further curated into 33 chromosome-scale scaffolds in each strain, guided by Hi-C contact maps (**Table 2.1, Figure S2.1**).

Table 2.1 : *R. irregularis* homokaryotic strains genome assembly and annotation statistics

	DAOM197198	**A1**	**B3**	**C2**	**4401**
Chromosome-scale scaffolds					
Total size	147.2 Mb	147.1 Mb	146.8 Mb	161.9 Mb	146.8 Mb
# of scaffolds	33	33	33	33	33
BUSCO completeness	96%	96%	96%	96.5%	96.5%
Repeat content	49.05%	49.00%	50%	54.87%	48.07%
Number of genes	26,634	23,988	23,988	25,634	24,069
GC content	27.82%	27.78%	27.81%	28.23%	27.81%
Unplaced contigs					
Total size	2.6 Mb	0.5 Mb	1.3 Mb	0.4 Mb	0.9 Mb
# of contigs	104	29	42	12	25
Assembly N/L50	29/31.4 Kb	7/25.5 Kb	12/36.5 Kb	4/50.8 Kb	7/45.7 Kb

When combining the five strains, we found telomeres on both ends for 32 of the 33 chromosomes (only chromosome 2 has telomeres on only one end in all strains). Genome sizes are very similar in the model strain DAOM197198, A1, B3 and 4401 at 146-147 Mb, while the C2 strain harbors a larger genome at 162 Mb, consistent with published flow cytometry analyses (Ropars *et al.,* 2016). For all five strains, the number of unplaced contigs is small (12-104 contigs at size 0.4-2.6 Mb) and all unplaced contigs are short indicating that the chromosome assemblies are near-complete (**Table 2.1, Figure S2.1**).

We identified an average of 76 families of repeats and known transposable elements (TE) encompassing on average 50.2% of the genome (**Figure S2.2**). When only TEs are considered, repeat density varies across chromosomes of different strains. For DAOM197198, A1 and B3 assemblies, chromosome 2 has the highest TE density (17.2%, 14.8% and 14.8% respectively), whereas the highest TE density is chromosome 33 for the C2 assembly (20.8%), and chromosome 23 for the 4401 assembly (15.6%). The chromosome with the lowest TE density is chromosome 20 for A1 and B3 (8.5% and 9.4%, respectively), chromosome 28 for DAOM197198 (9.2%), chromosome 3 for C2 (11.9%) and chromosome 8 for 4401 (7.3%).

Gene annotation identified between 23,693 to 26,820 genes, in line with past work based on fragmented datasets (Chen *et al.,* 2018b), and BUSCO completeness averages 96% (**Table 2.1**). Gene density varies among chromosomes, with chromosome 32 always having the highest density while chromosome 15 the lowest (**Figure S2.1**). A hallmark of AMF is that they carry divergent rRNA operons within their genome. In DAOM197198 one rRNA operon is found on chromosome 9, two operons co-locate on chromosomes 23 and 28, and five are present in chromosome 18 (**Figure S2.1)**. The putative AMF mating-type locus (Ropars *et al.,* 2016) is located on chromosome 11 in all strains (**Figure S2.1**).

Within species chromosome diversity in *R. irregularis*

Consistent with previous reports on fragmented assemblies (Chen *et al.*, 2018b), the *R. irregularis* gene content is divided into genes shared by all strains (core genes) and genes shared by only a few strains or strain-specific (dispensable) (**Figure S2.3a**), and on average only 55.8% of *R. irregularis* genes are core (Chen *et al.*, 2018b). Our analyses confirm that within species variability affects the number of rRNA operons (**Figure 2.1a**, **Table S2.1**) (Corradi *et al.*, 2007), and each strain carries a distinct abundance of protein domains (**Figure S2.3b**) (Chen *et al.*, 2018b).

Homologous chromosomes often differ in size between strains. For example, the largest chromosome of the 4401 strain is chromosome 8 at 6.3 MB in size but this chromosome is only 5.2 MB in DAOM197198 (**Figure 2.1b**). The chromosome with the putative MAT-locus (Chromosome 11) varies in size from a maximum of 5.6 Mb in C2 to a minimum of 5 Mb in 4401, even though they carry the same mating type (**Figure 2.1c**). We used a *k*-mer distance measure to assess relatedness of strains and their chromosomes. Overall, the strain DAOM197198 is related to the strain 4401, strains B3 and A1 are related, and C2 is the most distant (**Figure S2.4**).

However, this is not consistent across all chromosomes. For example, chromosome 8 shows higher levels of similarity between strain DAOM197198 and the strains B3 and A1 while chromosome 25 shows higher levels of similarity between strain 4401 and the strains B3 and A1 (**Figure S2.4**). These results support past reports of phylogenetic incongruence and inter-strain genetic exchange in *R. irregularis* (Riley *et al.*, 2014; Chen *et al.*, 2018b)

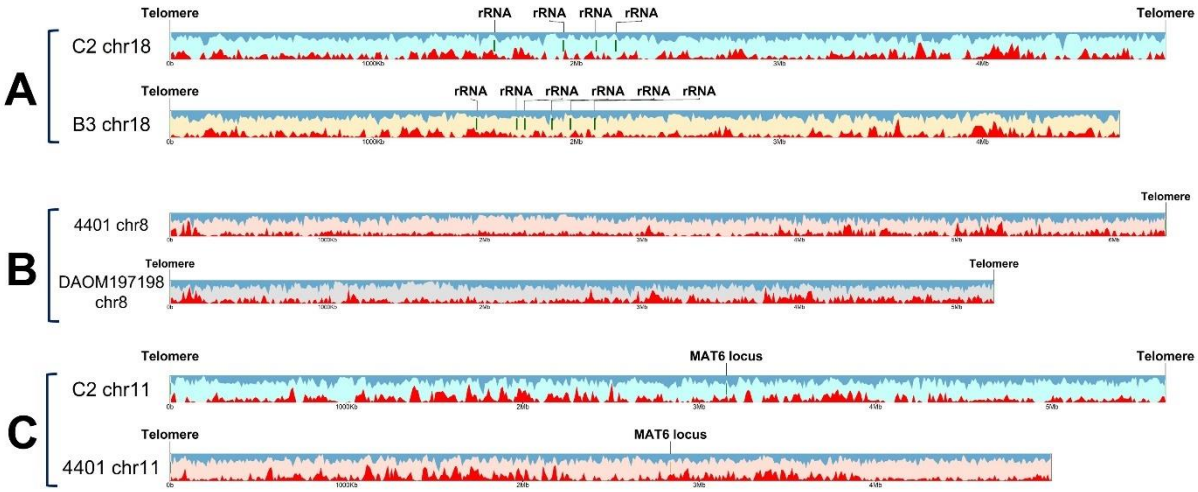

A
C2 chr18
Telomere
rRNA
rRNA
rRNA
rRNA
Telomere
B3 chr18
Telomere
rRNA
rRNA
rRNA
rRNA
rRNA
rRNA
B
4401 chr8
Telomere
DAOM197198 chr8
Telomere
Telomere
C
C2 chr11
Telomere
MAT6 locus
Telomere
4401 chr11
Telomere
MAT6 locus

Figure 2.1: Examples of variability in the size, and gene content and density between the homologous chromosomes of *R. irregularis* strains analyzed in this study. Red density plot shows gene density, and blue density plot shows repeat density. The colors of karyoplots indicate different strains. **A) Homologous chromosomes contain different rRNA operon copy numbers.** Chromosome 18 of C2 contains only four operon copies, whereas strain B3 carries six copies on the same chromosome. **B) Homologous chromosomes vary greatly in size**. Size difference between chromosome 8 of 4401 strain and chromosome 8 of DAOM197198 strain is over 1.1 Mb. **C) Chromosomes carrying MAT locus also vary in size.** C2 and 4401 strains carry the same MAT type, MAT6. However, chromosome sizes still differ by 700 kb.

The chromosomes of Rhizophagus irregularis form a two-compartment genome

Hi-C sequencing also generated direct, quantitative evidence of the 3D nuclear organization in five *R. irregularis* strains. For the strains DAOM197198, A1 and C2, this revealed that chromosomes form a distinct 'checkerboard' pattern that defines the long-range interactions that form two nuclear A/B compartments in eukaryotes, including fungi (Fortin & Hansen, 2015; Kim *et al.*, 2017; Spielmann *et al.*, 2018; Winter *et al.*, 2018; Falk *et al.*, 2019) (**Figure 2.2**, **Figures S2.5-2.10**).

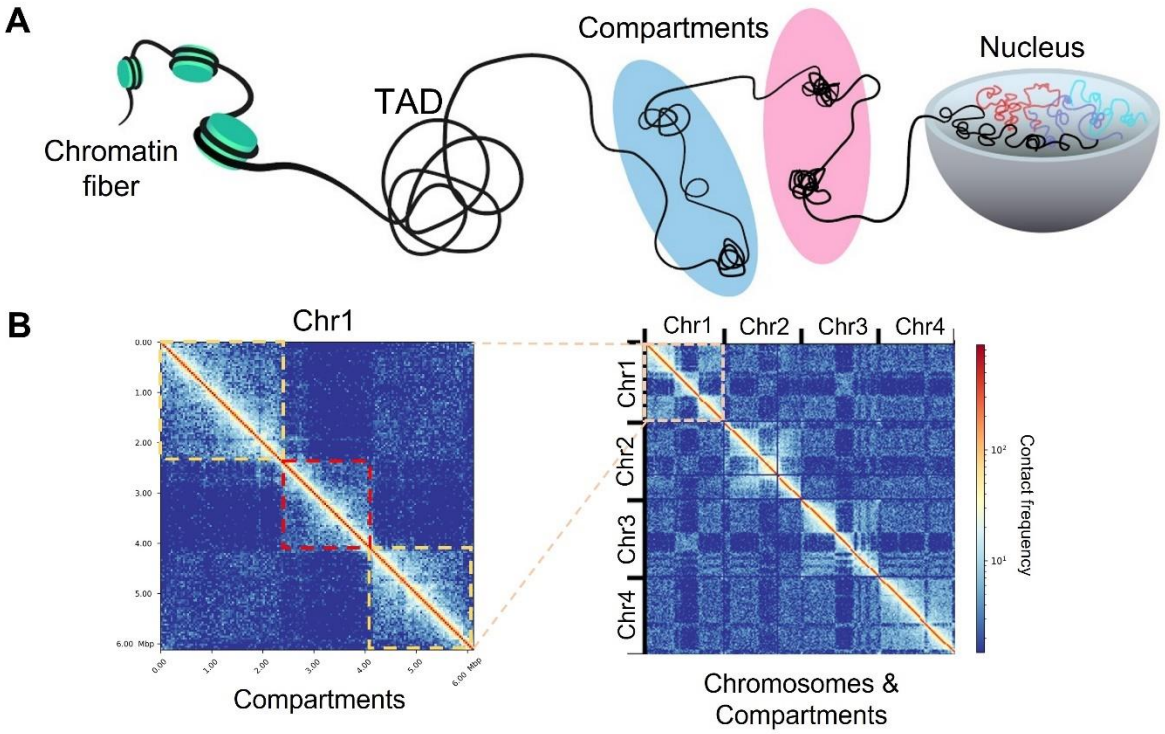

A
Chromatin fiber
TAD
Compartments
Nucleus
B
Chr1
Compartments
Chr1
Chr2
Chr3
Chr4
Contact frequency
Chromosomes & Compartments

Figure 2.2: Chromatin folding inside of the *R. irregularis* nucleus and visualization of intra- and inter-chromosomal physical interactions in Hi-C maps. A) Schematic showing the compaction of chromatin fibers inside of a nucleus. Chromatin fibers physically interact with each other and may fold into regions called compartments and topologically associated domains. **B) Hi-C contact maps showing the compartmentalization in the chromosome 1 of the *R. irregularis*.** In these heat maps, created to demonstrate contact frequencies throughout the genome, genome coordinates are represented on both axes. In an individual chromosome, regions that interact with each other result in an increase in contact frequencies and reveals the compartmentalized nature of the chromosome. These bright squares highlight increased contact frequency within and between chromosomes, are further analyzed to group them into euchromatin or heterochromatin compartments. Left: For chromosome l, the compartment shown in the red square belongs to compartment A, and is surrounded by two B-compartment regions shown in orange squares. Right: when the interactions of several chromosomes are analyzed, a "checkered" pattern indicates the genome arrangement of the euchromatin or heterochromatin compartments of *R. irregularis*. Telomeres are clearly visible in the Hi-C maps at the tip of each chromosome.

In the model strain DAOM197198, the A-compartment has a total size of 58.1 Mb while B-compartment is 73.5 Mb, and respectively each contains 96 and 83 TADs. Chromosome varies in the degree of compartmentalization – e.g. chromosome 2 has consistent swaps between A- and B-compartment, while the A-compartment dominates in the smallest chromosomes (chr30 – 33) **(Figure S2.6-2.10)**. Although *Rhizophagus irregularis* Hi-C contact maps did not show easily recognizable centromere hotspots found in rust fungi (Sperschneider *et al.*, 2021), these reveal telomere-to-telomere interactions that define all chromosomes boundaries and confirm our assembly, and inter-chromosomal physical connections between rRNA operons. Changes in Hi-C contacts between homologous chromosomes of different strains are also linked with the emergence of large strain-specific repeat and gene expansions (**Figure S2.11a**), or inversions (**Figure S2.11b,c,d**).

In the model strain DAOM197198, the A-compartment is more gene dense with 2.55 genes and 18.21 repeats per 10 Kb, while the B-compartment carries an average of 1.74 genes and 18.9 repeats per 10 Kb (**Figure 2.3a, Table S2.2)**. The A-compartment also codes for 90% of BUSCO proteins, and double the number of core genes (**Figure S2.12**), while B-compartment carries most dispensable genes and many highly expanded families – e.g. Tyrosine Kinases, the tetratricopeptide repeat Sel-1, the homodimerization BTB. The large high mobility box (HMG) gene family is, however, predominant in the B-compartment (**Figure 2.3b**). Remarkably, while the B-compartment is less gene-dense, it is enriched for predicted secreted proteins as well as for candidate apoplastic and cytoplasmic effectors and carries the cytoplasmic *R. irregularis* effector SP7 on chromosome 1(Kloppholz *et al.*, 2011).

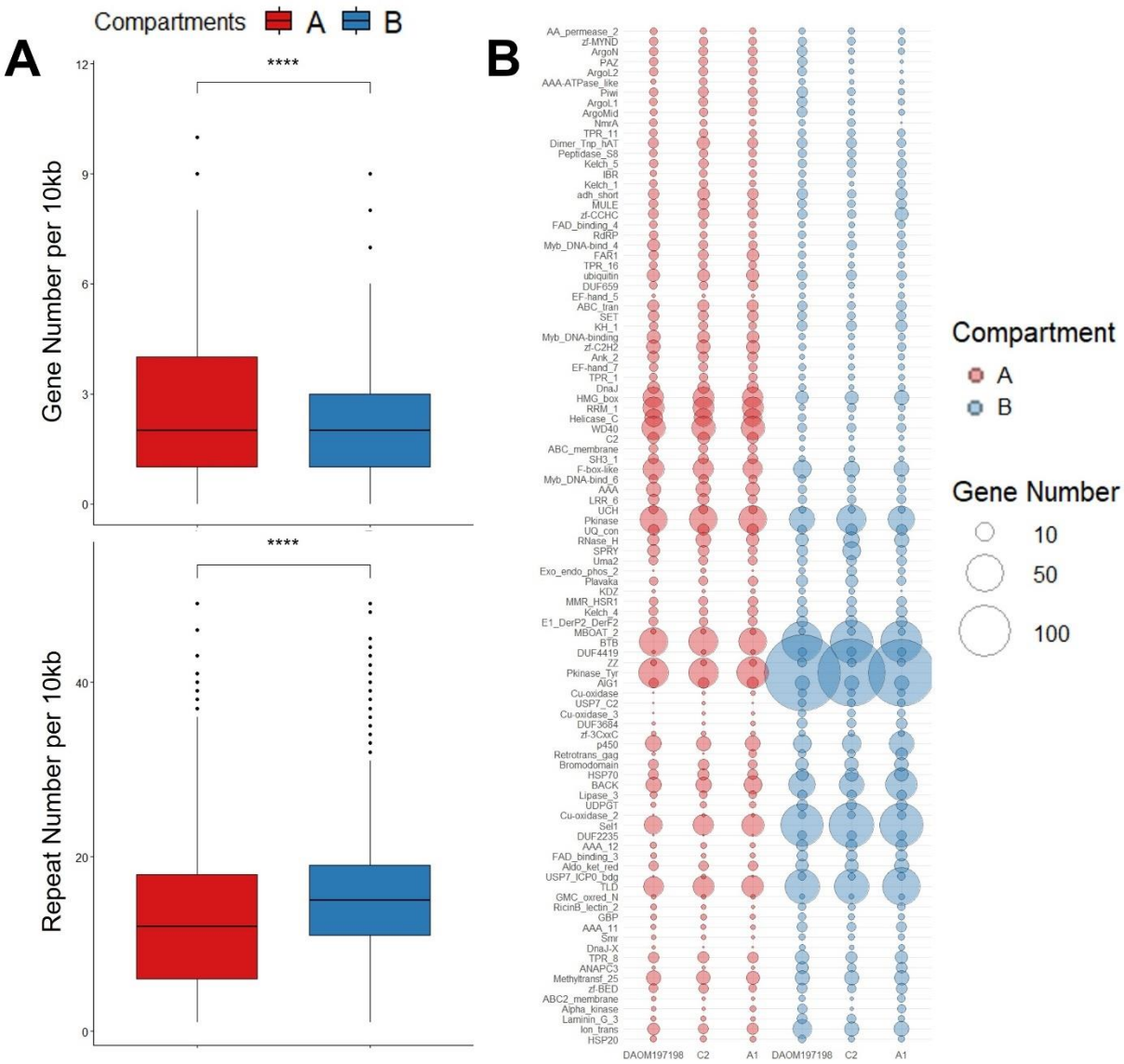

A
Compartments
A
B
****
Gene Number per 10kb
12
9
6
3
0
****
Repeat Number per 10kb
40
20
0
B
AA_permease_2
zf-MYND
ArgoN
PAZ
ArgoL2
AAA-ATPase_like
Piwi
ArgoL1
ArgoMid
NmrA
TPR_11
Dimer_Tnp_hAT
Peptidase_S8
Kelch_5
IBR
Kelch_1
adh_short
MULE
zf-CCHC
FAD_binding_4
RdRP
Myb_DNA-bind_4
FAR1
TPR_16
ubiquitin
DUF659
EF-hand_5
ABC_tran
SET
KH_1
Myb_DNA-binding
zf-C2H2
Ank_2
EF-hand_7
TPR_1
DnaJ
HMG_box
RRM_1
Helicase_C
WD40
C2
ABC_membrane
SH3_1
F-box-like
Myb_DNA-bind_6
AAA
LRR_6
UCH
Pkinase
UQ_con
RNase_H
SPRY
Uma2
Exo_endo_phos_2
Plavaka
KDZ
MMR_HSR1
Kelch_4
E1_DerP2_DerF2
MBOAT_2
BTB
DUF4419
ZZ
Pkinase_Tyr
AIG1
Cu-oxidase
USP7_C2
Cu-oxidase_3
DUF3684
zf-3CxxC
p450
Retrotrans_gag
Bromodomain
HSP70
BACK
Lipase_3
UDPGT
Cu-oxidase_2
Sel1
DUF2235
AAA_12
FAD_binding_3
Aldo_ket_red
USP7_ICP0_bdg
TLD
GMC_oxred_N
RicinB_lectin_2
GBP
AAA_11
Smr
DnaJ-X
TPR_8
ANAPC3
Methyltransf_25
zf-BED
ABC2_membrane
Alpha_kinase
Laminin_G_3
Ion_trans
HSP20
DAOM197198
C2
A1
DAOM197198
C2
A1
Compartment
A
B
Gene Number
10
50
100

Figure 2.3: A/B compartments have different gene (Gene/10Kb) and repeat (Repeat/10Kb) densities, and their genes contain different pfam domains. A) Upper panel: The Box-plot ranges from 0-10 and 0-9 for the A and B-compartments. Lower: The Box-plot ranges from 1-49 for both compartments. Box edges show the third quartile and first quartile in box plots, median is shown with a middle lane and the range of the data is indicated by the whiskers. Outliers are shown as dots. Asterisks indicate T-test statistical significance at $p < 0.0001$. **B) Gene numbers that carry specific Pfam domains also vary between compartments**. Circles highlight the total of number of genes carrying specific Pfam domains in located in compartment A (red circles) or B (blue circles), with the size of the circle being proportionate to the number of genes that carry that domain. Inter-strain variability in pfam domains is also evident within each compartment.

We hypothesized that the distinct gene and repeat densities result in different structural rearrangement rates between A/B compartments. We measured the degree of structural variation and sequence identity of each compartment, and found that the sequence alignments between strains in the B-compartment are shorter and have lower sequence identity than sequence alignments in the A-compartment. The B-compartment also contains more structural variation (**Table S2.3**, **Figure S2.13**).

Gene expression and methylation in chromosomal compartments

In other organisms, euchromatin (A compartments) is generally transcriptionally active while heterochromatin (B compartments) are repressed (Szabo *et al.*, 2019; Zheng & Xie, 2019; Jerković *et al.*, 2020). To test if this dichotomy also applies to AMF, we investigated gene expression between compartments in DAOM197108 using three high quality RNA-seq datasets obtained from intra-radical hyphae in symbiosis with *Allium*, arbuscocytes in symbiosis with *Medicago* and germinating spores (Kamel *et al.*, 2017b; Zeng *et al.*, 2018).

In *R. irregularis*, genes in the A compartment have, on average, significantly higher expression levels than those in the B compartment (**Figure 2.4a**). However, this compartment also has lower expression levels in the *in planta* conditions (*Allium*: 57.3 TPMs; *Medicago*: 56 TPMs) than in germinated spores (62.4 TPMs). Remarkably, the inverse pattern is observed for genes in the B-compartments, which have higher expression levels in the *in planta* conditions (*Allium*: 16.8 TPMs; *Medicago*: 19.5 TPMs) than in germinated spores (11.2 TPMs), and genes encoding secreted proteins are significantly up-regulated in both *in planta* conditions in the B-compartment, but not in the A-compartment (**Figure 2.4b**). We also aimed to investigate the expression of these regions within TAD, and found that genes tend to be co-regulated when they

are located in the same TAD in the A-compartment, while co-regulation within the same TAD was not observed for the B-compartment **(Figure S2.14)**.

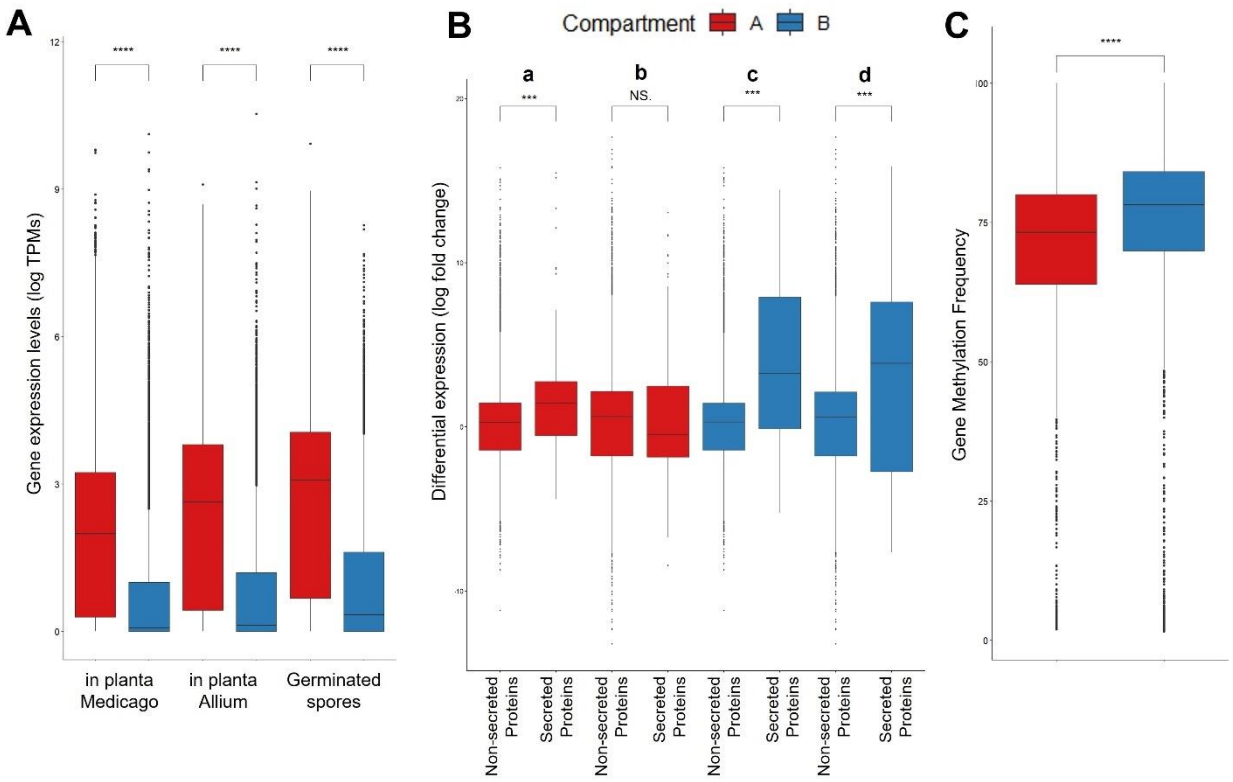

A
Gene expression levels (log TPMs)
****
****
****
in planta Medicago
in planta Allium
Germinated spores
B
Compartment
A
B
a
b
c
d
***
NS.
***
***
Differential expression (log fold change)
Non-secreted Proteins
Secreted Proteins
C
Gene Methylation Frequency
****

Figure 2.4: Gene expression and methylation analyses for the A/B compartments. **A) Genes in the A compartment show significantly higher expression levels than genes in the B-compartment in all three conditions.** Compartment A and B boxplot values range between 0-9.81 and 0-10.12 for in planta *Medicago* condition, between 0-9.09 and 0-10.53 for in planta *Allium* condition and 0-9.93 and 0-8.27 in germinated spores, respectively. **B) Genes encoding secreted proteins are up-regulated in both in planta conditions in the B-compartment, but not in the A-compartment. a)** The A-compartment shows significant upregulation of secreted proteins *in planta* Allium condition when compared to germinated spores. For A-compartment, non-secreted protein boxplot values range between -11.17 and 15.78, while secreted protein boxplot values range between -4.40 and 15.47 when in planta *Allium* expression is compared to the expression of germinated spores. **b)** The A-compartment does not show a significant secreted protein expression change when *in planta* Medicago expressions are compared to the expression of germinated spores. For A-compartment, non-secreted protein boxplot values range between -13.22 and 17.66, while secreted protein boxplot values range between -8.44 and 13.06 when in planta *Medicago* expression is compared to the expression in germinated spores. **c)** The B-compartment shows significant upregulation of secreted proteins *in planta* Allium condition when compared to germinated spores. For B-compartment, non-secreted protein boxplot values range between -11.17 and 15.08, while secreted protein boxplot values range between -5.25 and 14.46 when in planta *Allium* expression is compared to the expression in germinated spores. **d)** The B-compartment B shows significant upregulation of secreted proteins *in planta* Medicago condition when compared to germinated spores. For B-compartment, non-secreted protein boxplot values range between -13.22 and 17.66, while secreted protein boxplot values range between -7.65 and 15.86 when in planta *Medicago* expression is compared to the expression in germinated spores. **C) Gene median methylation frequencies of all methylated genes (methylation frequency median > 0) that are located in A/B compartments.** The A-compartment displays significantly lower gene median methylation frequencies than the B compartment. A compartment boxplot ranges between 2 and 100, whereas B compartment boxplot ranges between 1.6 and 100. Asterisks indicate T-test statistical significance at $p < 0.001$ (***) and < 0.0001 (****). NS = not significant. Box edges show the third quartile and first quartile in box plots, median is shown with a middle lane and the range of the data is indicated by the whiskers. Outliers are shown as dots. TPM = Transcripts Per Million reads.

We used ONT sequencing to call methylated sites and test if the difference in the A/B compartment gene expression resulted from different methylation patterns (Lea *et al.*, 2018; Dallaire, 2021). In DAOM197198 strain, CG dinucleotides are highly enriched for 5mC methylation – i.e. 11.88% for CG compared to 0.72 to 3.95% for other dinucleotides contexts (CA, CC, CT) – 6mA methylation is very low regardless of dinucleotide contexts (AA, AC, AG, AT: 0.01% to 0.06). This indicates that 5mC methylation in the CG dinucleotide context is the primary DNA methylation process active in this species.

Overall, the CpG sites are significantly more methylated in the A-compartment, particularly for highly methylated sites (methylation frequency > 80%) (**Figure S2.13a)**. Within the B-compartment, 22.5% of all protein encoding genes are highly methylated compared to 18.4% in the A-compartment (**Figure 2.4c**). When TEs are considered, this is inversed with 67% of TEs families (24 out of 35) being highly methylated in the A-compartment compared to only 25% for the B -compartment (9 out of 36) (**Figure S2.13b**). Among TEs that are compartment specific - e.g Ginger DNA 2, DNA Zisupton and LINE Penelope are only found in compartment B and the LINE -Tad, and LINE R1-LOA carried by the A-compartment – only those located in the A-compartment are methylated.

Discussion

A chromosome-level view of a model AMF pangenome

The present study confirms that the combined number of genes within the model AMF *R. irregularis* far exceeds that found in individual strains (Chen *et al.*, 2018b; Mathieu *et al.*, 2018; Reinhardt *et al.*, 2021). It also revealed that strains in this model species carry 33 homologous chromosomes (possibly the largest number identified in a fungus) and within-species variability in chromosome size and epigenetics (folding and methylation).

Size difference among homologous chromosomes affects all strains, regardless of their phylogenetic clustering(Chen *et al.*, 2018b; Savary *et al.*, 2018; Kokkoris *et al.*, 2021) or MAT-locus identity – e.g. strains C2 and 4401 carry the MAT-locus-6 (as defined by(Ropars *et al.*, 2016)) on chromosome 11 but this chromosome is 640 Kb larger in C2 compared to 4401. Despite this variability, all strains analysed carry the same number of homologous chromosomes, and future studies may reveal if this characteristic defines species boundaries in this taxonomically challenging taxon (Bruns *et al.*, 2018).

Chromosome compartments dictate AMF genome biology and evolution

Hi-C data revealed that AMF chromosomes separate into locations with euchromatin (transcriptionally active A-compartment) and heterochromatin (transcriptionally less active B-compartments) that regulate both the expression and evolution of distinct molecular functions. In the A-compartment, higher gene density and repeat methylation indicates a tighter control of repeat expansions that could be deleterious for a region that contain most core genes. In contrast, the B-compartment experiences higher gene and lower repeat methylation rates, which explains its reduced gene expression levels and higher rearrangement rates.

It was proposed that AMF and notorious plant pathogens (e.g. *Fusarium oxysporum*, *Verticillium dahlia*; *Zymoseptoria tritici*) evolved similar genomic strategies to cope with their evolving plant hosts and competitors (Reinhardt *et al.*, 2021). The A/B compartments are reminiscent of the "two-speed genome architecture" reported in some fungal plant pathogens, where highly repetitive and rapidly rearranging genomic regions that express secreted proteins (i.e. analogous to B-compartments) separate from gene-rich locations that carry most core genes (i.e. analogous to A-compartments) (Mathieu *et al.*, 2018). Another similarity with plant pathogens is the presence in *R. irregularis* strains of many dispensable and lineage-specific genes. Despite this genomic resemblance, however, our analyses did not reveal the presence in *R. irregularis* of dispensable chromosomes that drive the adaptation of fungal plant pathogens to changing environments (Garmaroodi & Taga, 2007; Vlaardingerbroek *et al.*, 2016; Bertazzoni *et al.*, 2018).

A role of a two-compartment genome in plant colonization?

We showed that the A/B compartments have distinct epigenetic signatures, and in line with mammalian studies, the A-compartment in *R. irregularis* is transcriptionally active whereas the B-compartment is repressed. However, we also found that this does not hold true for genes encoding secreted proteins and candidate effectors. During two *in planta* conditions, we observed increased up-regulation of genes encoding secreted proteins in the B-compartment, but not in the A-compartment. This suggests that the repressed state of the B-compartment might be relaxed during plant colonization, possibly induced through signals from the plants (Plett & Martin, 2012). In the fungal pathogen *Leptosphaeria maculans* epigenetic control mechanisms lead to effector gene regulation (Soyer *et al.*, 2014). As our Hi-C data derives from extraradical mycelium, acquiring similar data *in planta* may show that intra-radical mycelium carries distinct

chromosome conformations that lift repression for the *R. irregularis* B-compartment (Frantzeskakis *et al.*, 2019; Torres *et al.*, 2020).

Conclusions

Combining long-reads with Hi-C sequencing demonstrates that model arbuscular mycorrhizal symbionts all carry 33 chromosomes with contents and sizes that varies significantly among strains. This supports the notion that AMF have pangenomes and, more generally, that conspecific strains should never be assumed to have identical genomes and genes (Malar C *et al.*, 2021b).

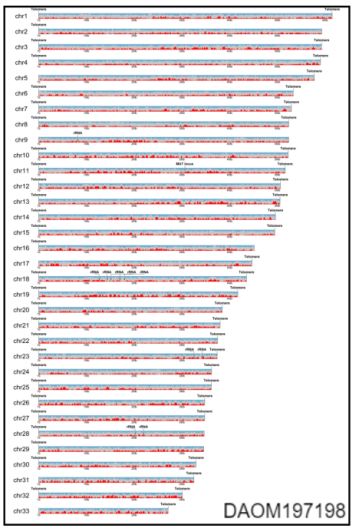
Our work also uncovered a higher-order genome organization that governs AMF genome biology and evolution. Dispensable genes and the most expressed and regulated secreted proteins - i.e. those potentially involved in the molecular dialogue with the plant hosts - locate primarily in the rapidly evolving B-compartments, while highly expressed core genes and tightly regulated TADs are mostly restricted within the slowly evolving A-compartment. In other organisms, the A/B compartments change depending on the cell life-stage (Falk et al., 2019). Within this context, our findings raise the intriguing possibility that similar changes also occur depending on the fungal and plant symbiotic status, and open avenues to identify the epigenetic mechanisms that generate and modify chromosome compartments during host-microbe interactions.

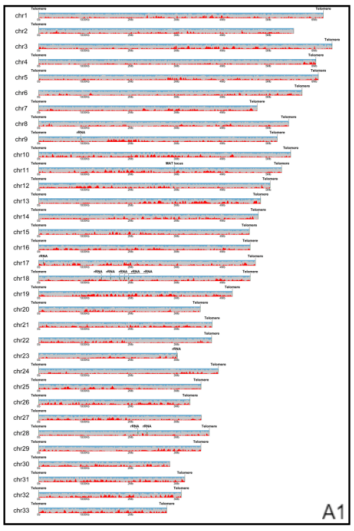
It will be now interesting to examine how this work extrapolates to AMF dikaryons, in particular how these strains compartmentalize their co-existing parental genomes and if these also significantly differ in gene content and size. Phasing parental genomes with high-quality Hi-C data is a requirement to fully address these questions, and determine how co-existing genomes regulate nuclear dynamics in these strains (Kokkoris *et al.*, 2021; Serghi *et al.*, 2021).

Acknowledgments

We thank Benjamin Schwessinger and Daniel Croll for their helpful comments on an earlier version of this manuscript, and Christophe Roux for providing information about publicly available, high-quality RNA-seq data. We are also grateful to Jon Hultqvist and Andrew Roger for providing training on Nanopore data acquisition and analysis. Our research is funded by the Discovery program of the Natural Sciences and Engineering Research Council (RGPIN2020-05643) and Discovery Accelerator Supplements Program (RGPAS-2020-00033). N.C. is a University of Ottawa Research Chair in Microbial Genomics. J.S. is supported by an Australian Research Council (ARC) Discovery Early Career Researcher Award (DE190100066), and E.C. and W.I. were supported by JSPS Postdoctoral Fellowships for Research in Japan and JSPS KAKENHI Grant Number 19F19089.

Supplemental figures and tables legends





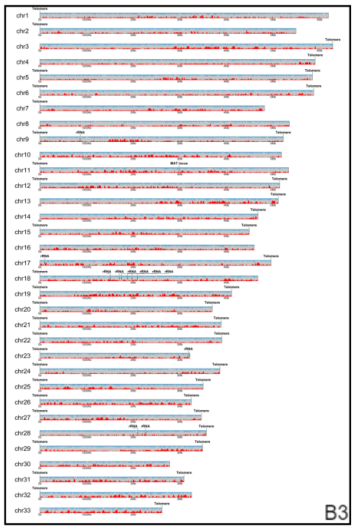


Figure S2.1: *R. irregularis* DAOM197198, C2, A1, B3 and 4401 strain chromosome karyoplots. Red density plots represent gene densities and blue density plots represent repeat densities. Telomeres and rRNA operon sites are marked.

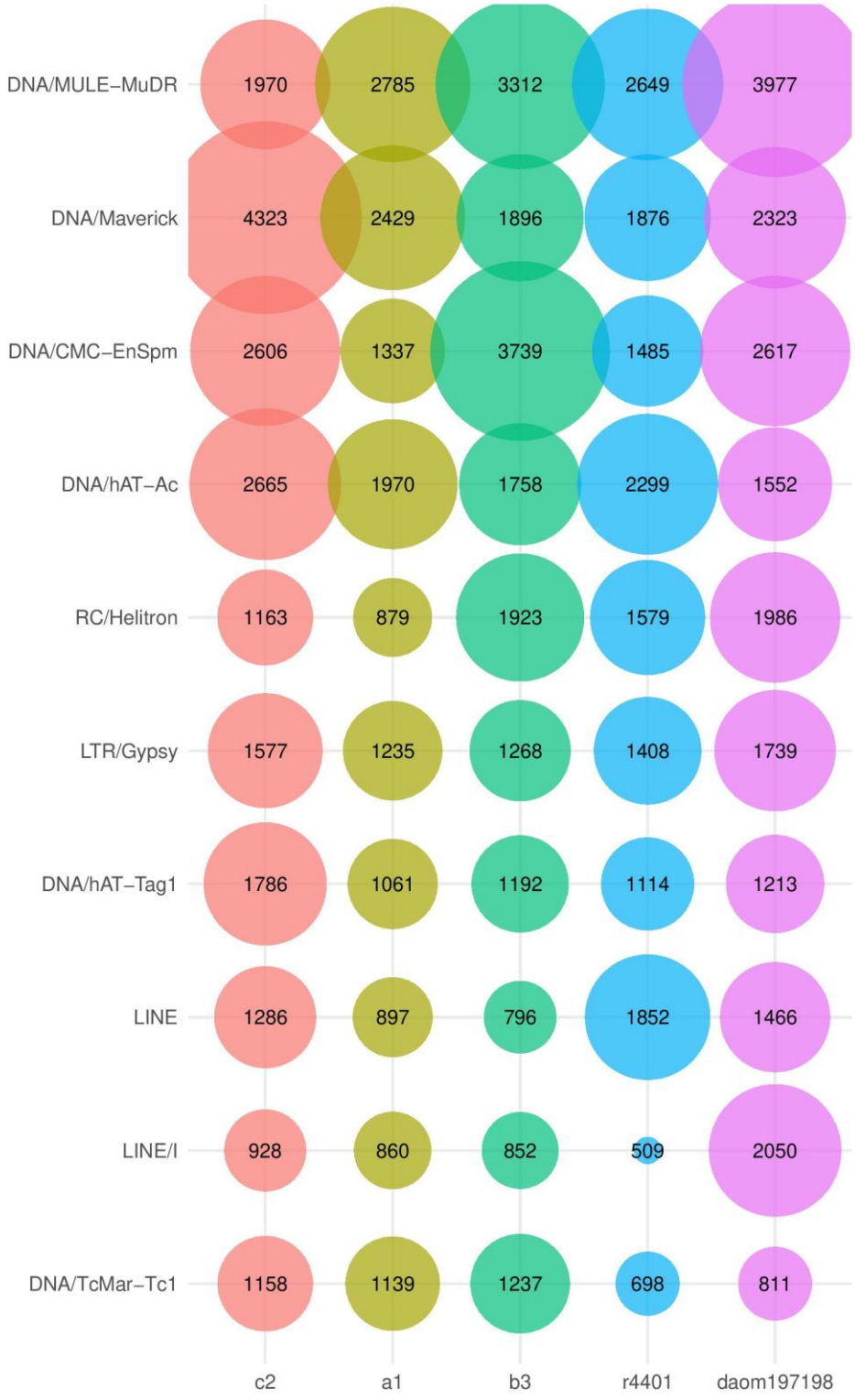

DNA/MULE−MuDR
1970
2785
3312
2649
3977
DNA/Maverick
4323
2429
1896
1876
2323
DNA/CMC−EnSpm
2606
1337
3739
1485
2617
DNA/hAT−Ac
2665
1970
1758
2299
1552
RC/Helitron
1163
879
1923
1579
1986
LTR/Gypsy
1577
1235
1268
1408
1739
DNA/hAT−Tag1
1786
1061
1192
1114
1213
LINE
1286
897
796
1852
1466
LINE/I
928
860
852
509
2050
DNA/TcMar−Tc1
1158
1139
1237
698
811
c2
a1
b3
r4401
daom197198

Figure S2.2: Bubble plot showing 10 most abundant transposable elements and their distribution in five assemblies of *R. irregularis* strains. Each color represents a different *R. irregularis* strain.

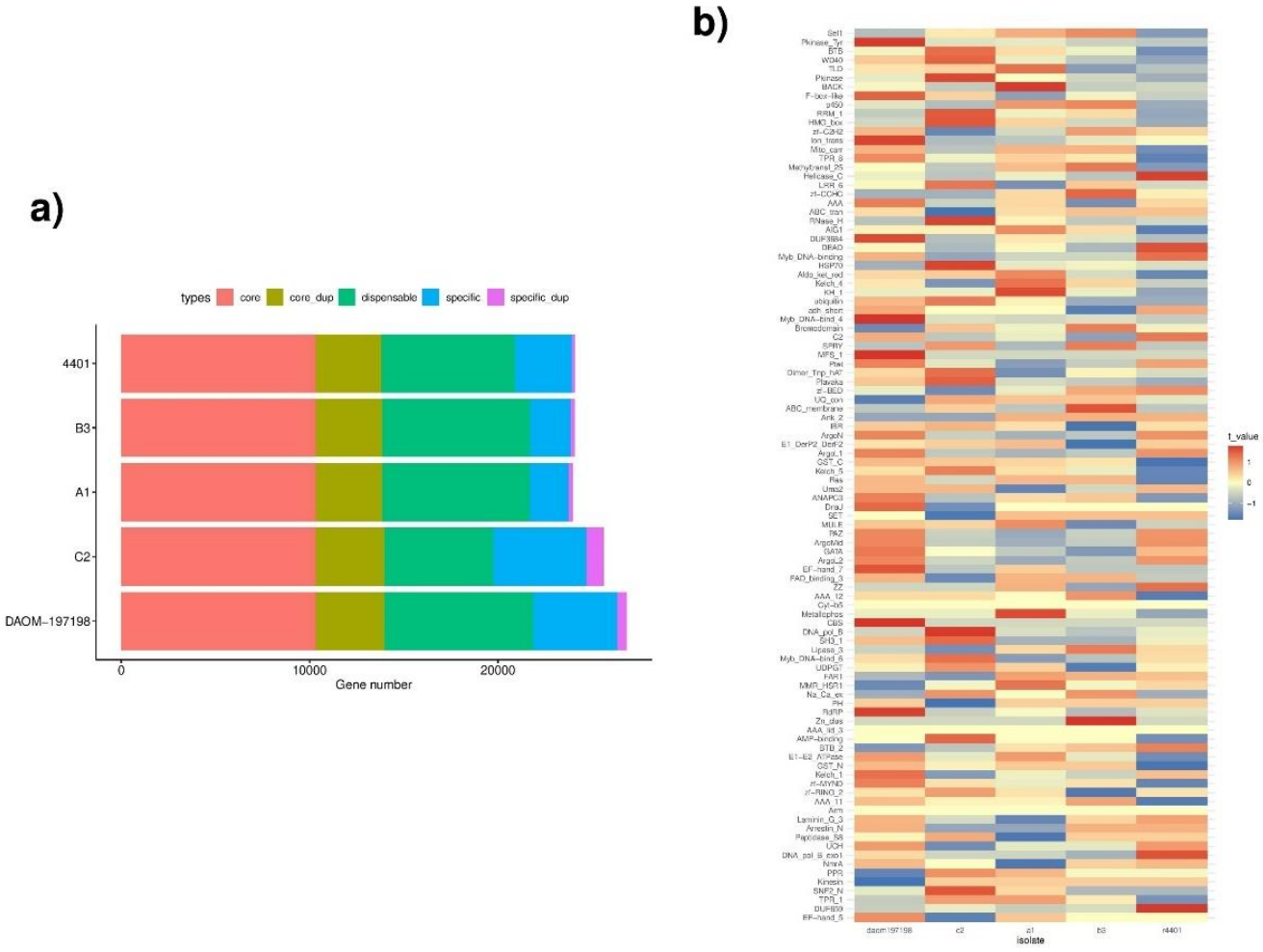

a)
types
core
core_dup
dispensable
specific
specific_dup
4401
B3
A1
C2
DAOM−197198
0
10000
20000
Gene number
b)
t_value
1
0
−1
daom197198
c2
a1
b3
n4401
isolate

Figure S2.3: Gene orthologies and Pfam domain numbers in five strains of *R. irregularis*. **a)** Colors represent the number of genes in each FastOrtho category. Single genes shared between all strains are considered core genes (red). If there are duplications present, they are core duplicates (yellow). Dispensable genes have orthologs in other strains, but not all. Specific genes are strain-specific, and if they are duplicated, they are found in specific duplicated group. **b)** The most abundant Pfam domains and their abundance in all strains. The abundance is transformed into t-values for each domain.

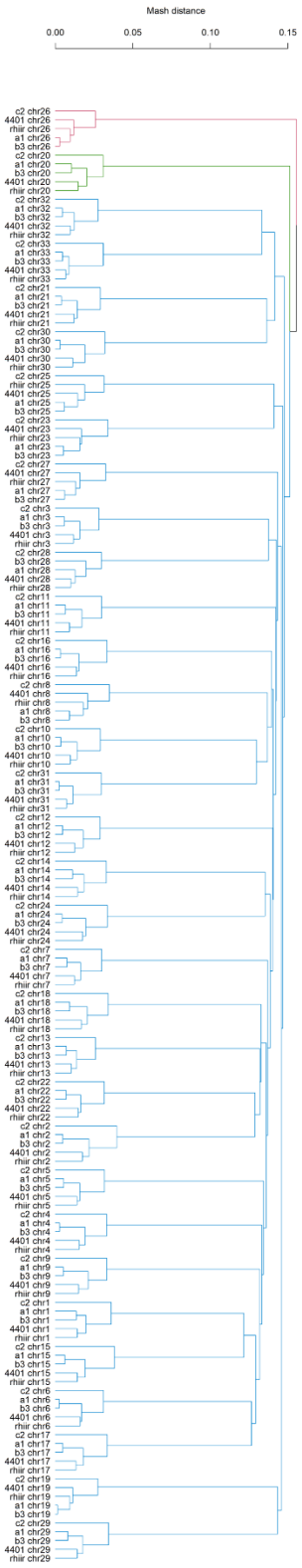
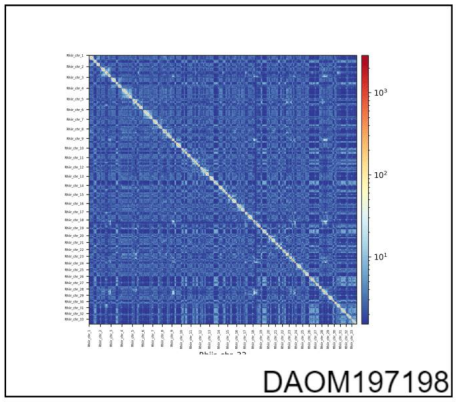
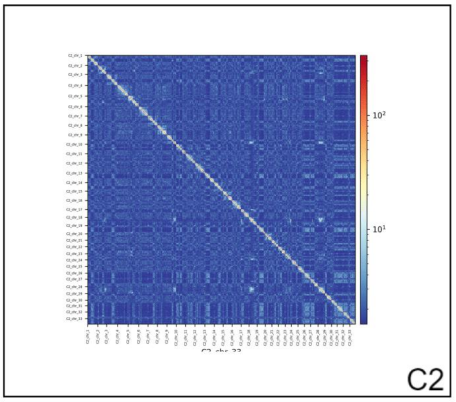
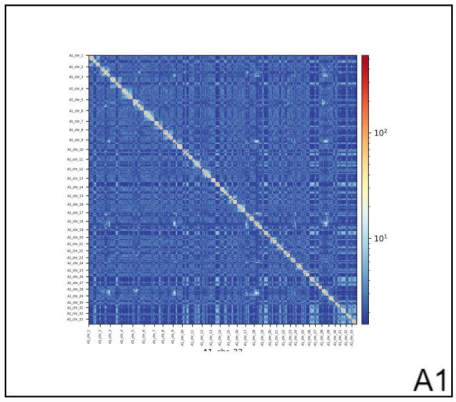
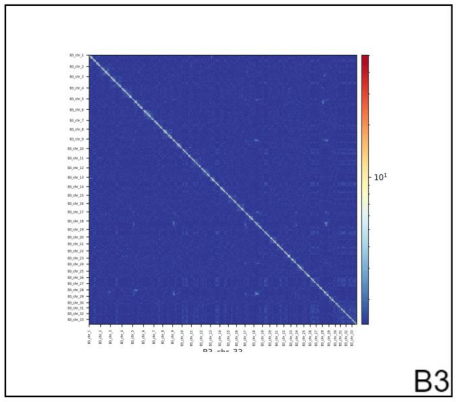

Mash distance
0.00
0.05
0.10
0.15
c2 chr26
4401 chr26
rhiir chr26
a1 chr26
b3 chr26
c2 chr20
a1 chr20
b3 chr20
4401 chr20
rhiir chr20
c2 chr32
a1 chr32
b3 chr32
4401 chr32
rhiir chr32
c2 chr33
a1 chr33
b3 chr33
4401 chr33
rhiir chr33
c2 chr21
a1 chr21
b3 chr21
4401 chr21
rhiir chr21
c2 chr30
a1 chr30
b3 chr30
4401 chr30
rhiir chr30
c2 chr25
rhiir chr25
4401 chr25
a1 chr25
b3 chr25
c2 chr23
4401 chr23
rhiir chr23
a1 chr23
b3 chr23
c2 chr27
4401 chr27
rhiir chr27
a1 chr27
b3 chr27
c2 chr3
a1 chr3
b3 chr3
4401 chr3
rhiir chr3
c2 chr28
b3 chr28
a1 chr28
4401 chr28
rhiir chr28
c2 chr11
a1 chr11
b3 chr11
4401 chr11
rhiir chr11
c2 chr16
a1 chr16
b3 chr16
4401 chr16
rhiir chr16
c2 chr8
4401 chr8
rhiir chr8
a1 chr8
b3 chr8
c2 chr10
a1 chr10
b3 chr10
4401 chr10
rhiir chr10
c2 chr31
a1 chr31
b3 chr31
4401 chr31
rhiir chr31
c2 chr12
a1 chr12
b3 chr12
4401 chr12
rhiir chr12
c2 chr14
a1 chr14
b3 chr14
4401 chr14
rhiir chr14
c2 chr24
a1 chr24
b3 chr24
4401 chr24
rhiir chr24
c2 chr7
a1 chr7
b3 chr7
4401 chr7
rhiir chr7
c2 chr18
a1 chr18
b3 chr18
4401 chr18
rhiir chr18
c2 chr13
a1 chr13
b3 chr13
4401 chr13
rhiir chr13
c2 chr22
a1 chr22
b3 chr22
4401 chr22
rhiir chr22
c2 chr2
a1 chr2
b3 chr2
4401 chr2
rhiir chr2
c2 chr5
a1 chr5
b3 chr5
4401 chr5
rhiir chr5
c2 chr4
a1 chr4
b3 chr4
4401 chr4
rhiir chr4
c2 chr9
a1 chr9
b3 chr9
4401 chr9
rhiir chr9
c2 chr1
a1 chr1
b3 chr1
4401 chr1
rhiir chr1
c2 chr15
a1 chr15
b3 chr15
4401 chr15
rhiir chr15
c2 chr6
a1 chr6
b3 chr6
4401 chr6
rhiir chr6
c2 chr17
a1 chr17
b3 chr17
4401 chr17
rhiir chr17
c2 chr19
4401 chr19
rhiir chr19
a1 chr19
b3 chr19
c2 chr29
a1 chr29
b3 chr29
4401 chr29
rhiir chr29

Figure S2.4: Hierarchical clustering of *k*-mer distance estimations of chromosomes.


DAOM197198


C2


A1


B3

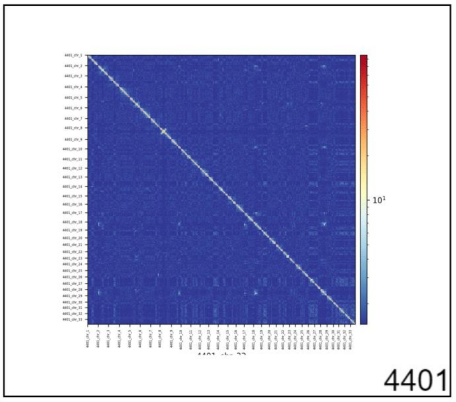
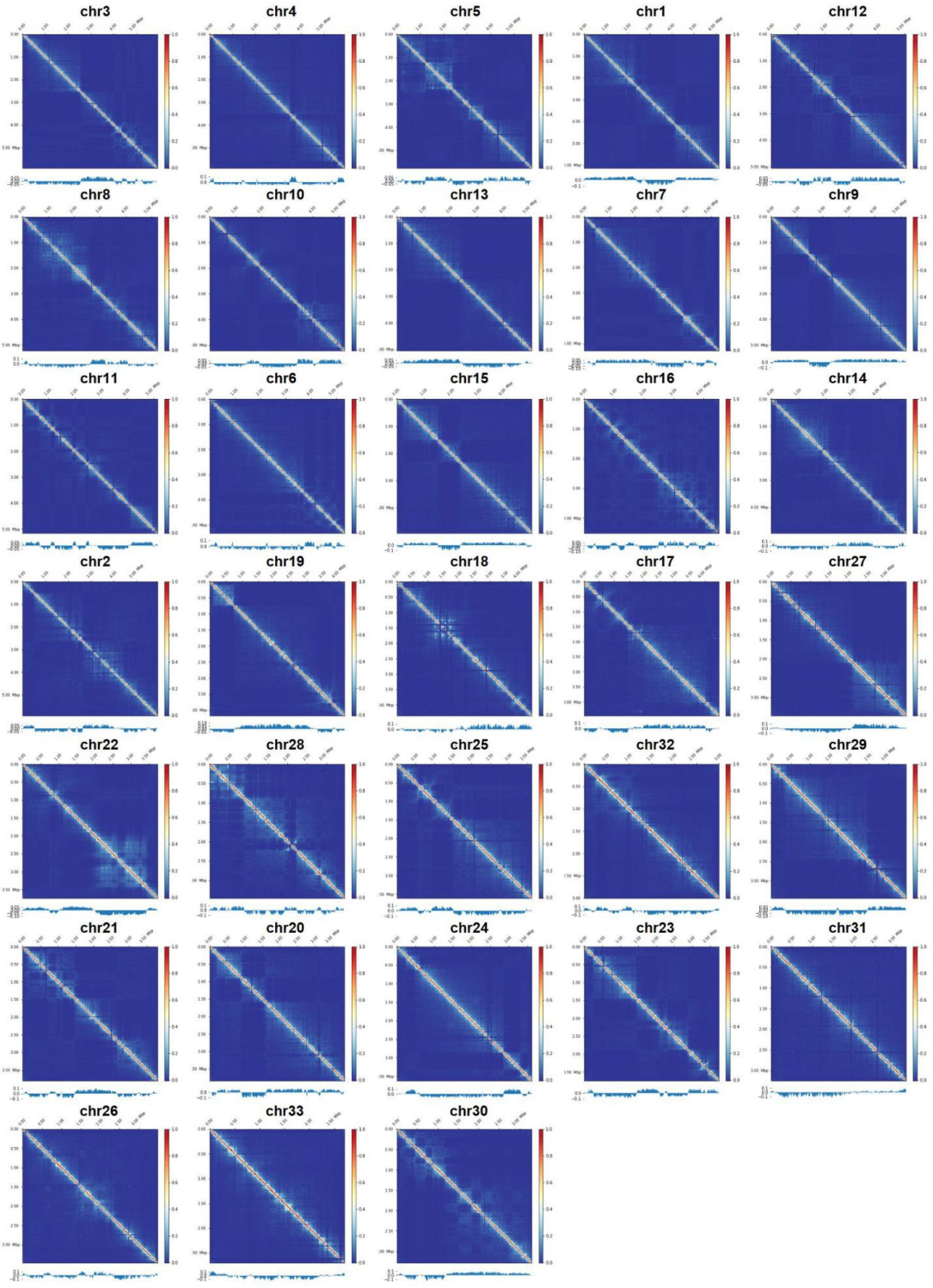
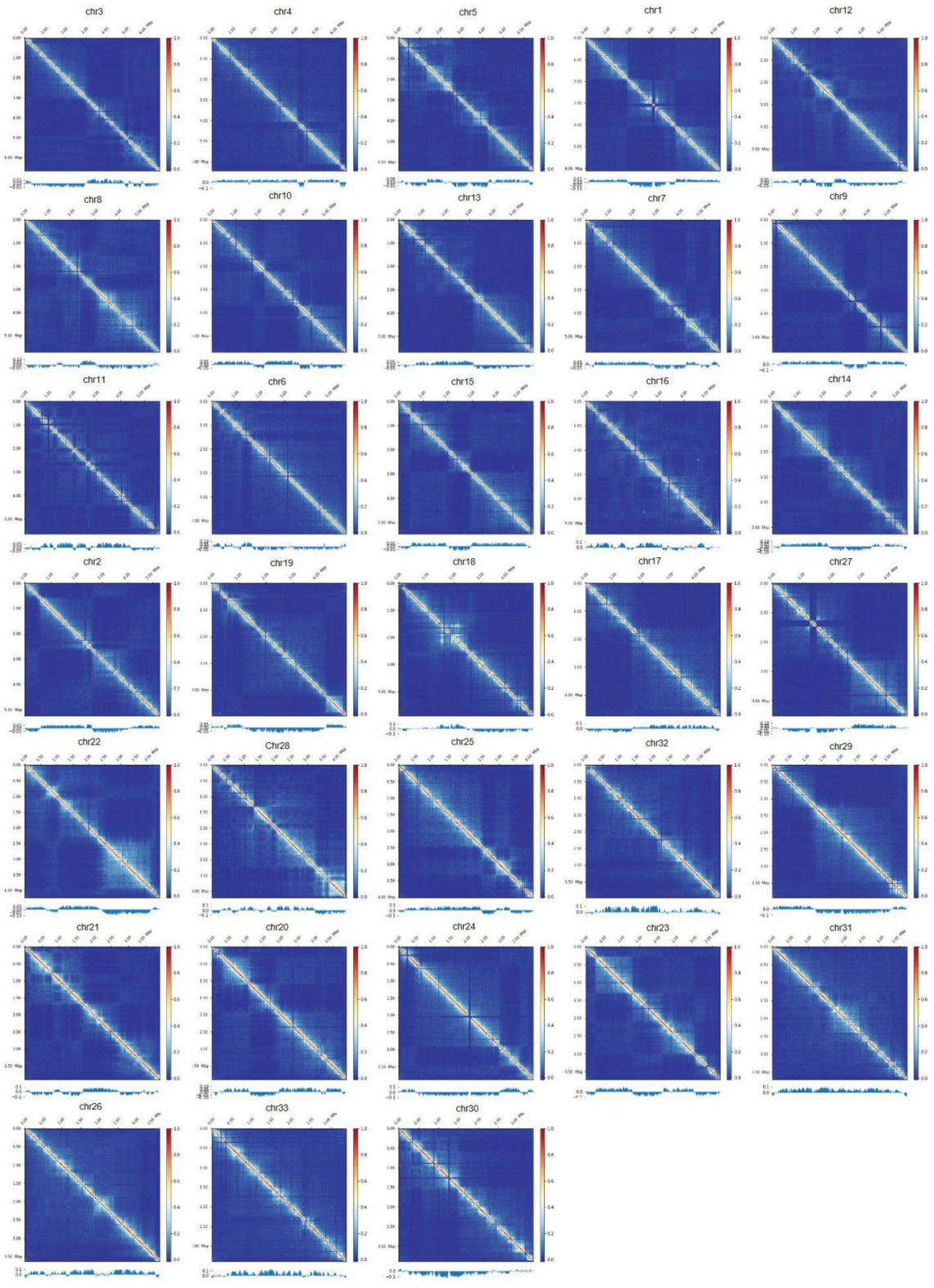
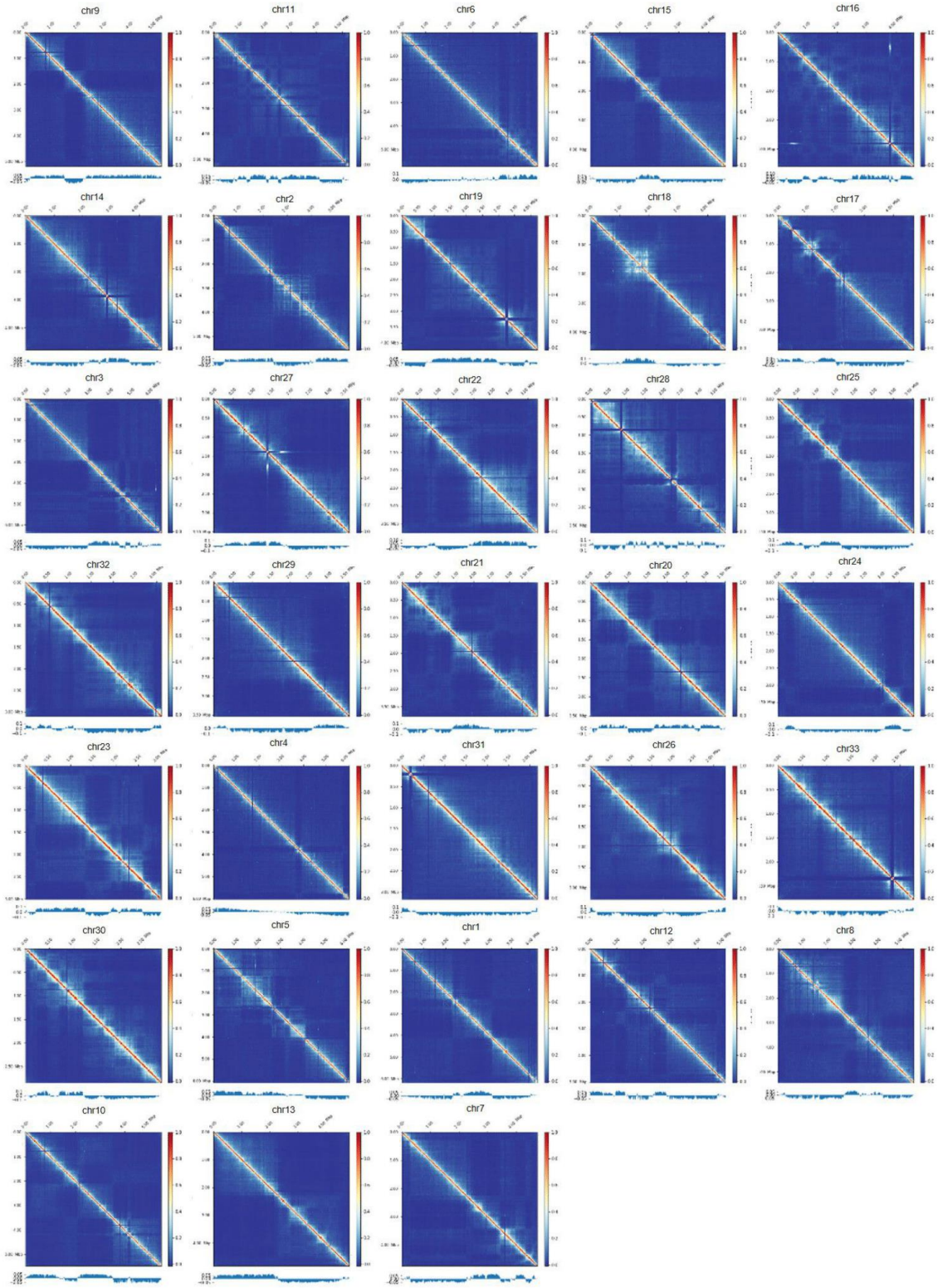
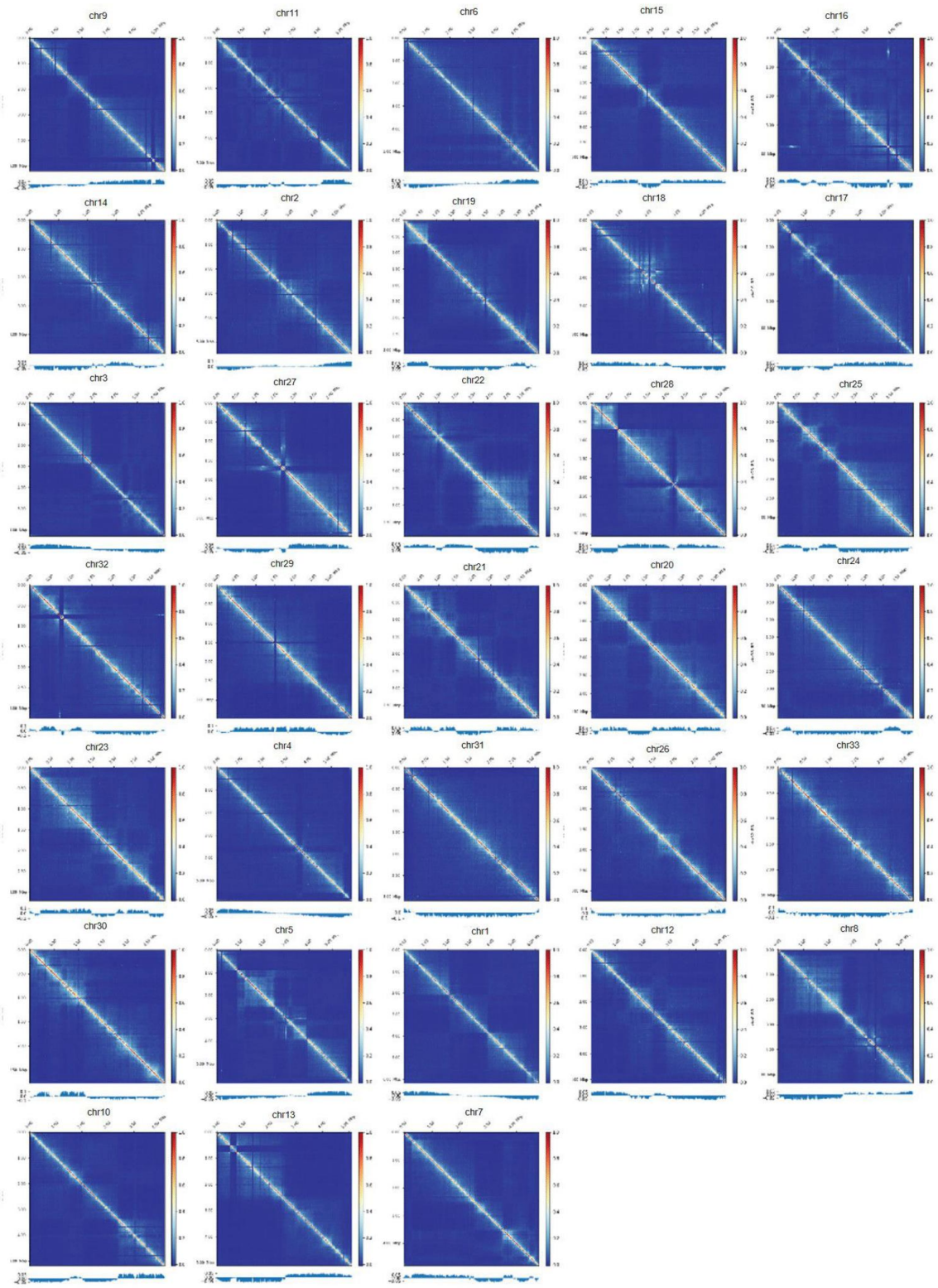

4401

Figure S2.5: Heat map showing Hi-C contact frequencies between 33 chromosomes for strains DAOM197198, C2, A1, B3 and 4401. The heatmaps were made using Hi-C Explorer hicPlotMatrix. Bright areas represent chromosomal regions that are in close proximity, found in the same compartment. Dark areas represent two chromosomal regions that are physically distant from each other. Checkered pattern represents a two-compartment genome structure.


chr3
chr4
chr5
chr1
chr12
chr8
chr10
chr13
chr7
chr9
chr11
chr6
chr15
chr16
chr14
chr2
chr19
chr18
chr17
chr27
chr22
chr28
chr25
chr32
chr29
chr21
chr20
chr24
chr23
chr31
chr26
chr33
chr30


chr3
chr4
chr5
chr1
chr12
chr8
chr10
chr13
chr7
chr9
chr11
chr6
chr15
chr16
chr14
chr2
chr19
chr18
chr17
chr27
chr22
Chr28
chr25
chr32
chr29
chr21
chr20
chr24
chr23
chr31
chr26
chr33
chr30


chr9
chr11
chr6
chr15
chr16
chr14
chr2
chr19
chr18
chr17
chr3
chr27
chr22
chr28
chr25
chr32
chr29
chr21
chr20
chr24
chr23
chr4
chr31
chr26
chr33
chr30
chr5
chr1
chr12
chr8
chr10
chr13
chr7


chr9
chr11
chr6
chr15
chr16
chr14
chr2
chr19
chr18
chr17
chr3
chr27
chr22
chr28
chr25
chr32
chr29
chr21
chr20
chr24
chr23
chr4
chr31
chr26
chr33
chr30
chr5
chr1
chr12
chr8
chr10
chr13
chr7

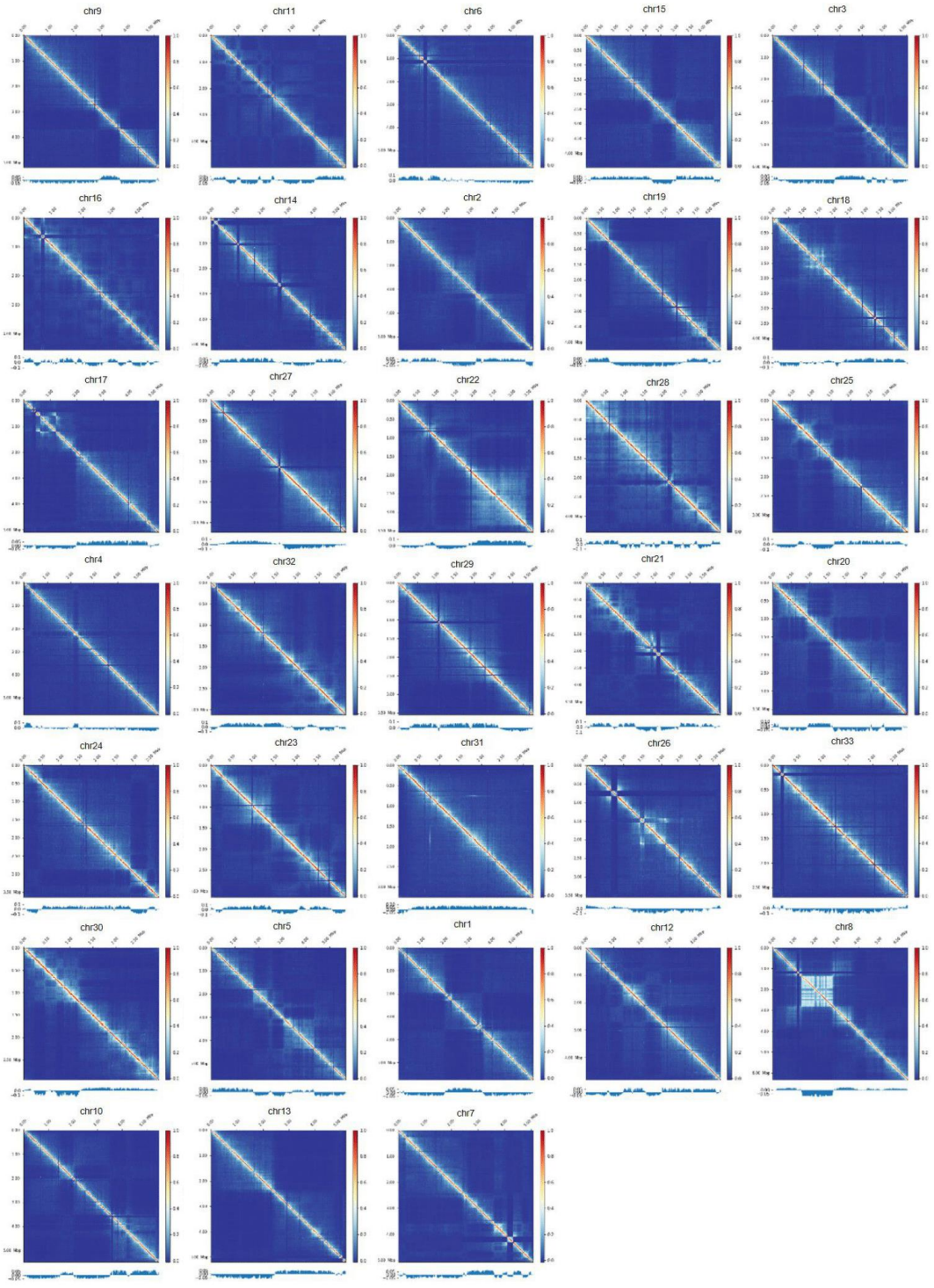
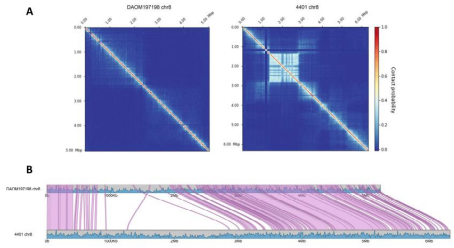
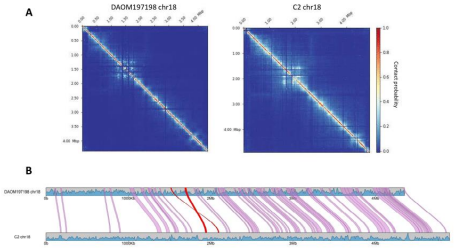

chr9
chr11
chr6
chr15
chr3
chr16
chr14
chr2
chr19
chr18
chr17
chr27
chr22
chr28
chr25
chr4
chr32
chr29
chr21
chr20
chr24
chr23
chr31
chr26
chr33
chr30
chr5
chr1
chr12
chr8
chr10
chr13
chr7

Figure S2.6-2.10: Heat map showing Hi-C contact probabilities of individual chromosomes of DAOM197198, C2, A1, B3 and 4401 strains analysed in this study. The heatmaps were made using Hi-C Explorer hicPlotMatrix. Bright areas represent chromosomal regions that are in close proximity, found in the same compartment. Dark areas represent two chromosomal regions that are physically distant from each other. Checkered pattern represents a two-compartment genome structure. The PCA plots at the bottom were calculated using Hi-C Explorer hicPCA. Above and below of x-axis represent two different compartments.

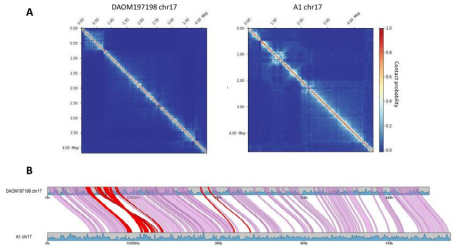
a)


A
DAOM197198 chr8
4401 chr8
Contact probability
B
DAOM197198 chr8
4401 chr8

b)


A
DAOM197198 chr18
C2 chr18
Contact probability
B
DAOM197198 chr18
C2 chr18

c)


A
DAOM197198 chr17
A1 chr17
Contact probability
B
DAOM197198 chr17
A1 chr17

d)

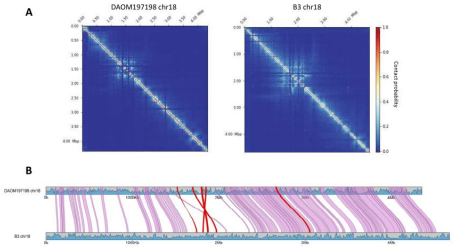

A
DAOM197198 chr18
B3 chr18
Contact probability
B
DAOM197198 chr18
B3 chr18

Figure S2.11: Strain-specific chromosomal differentiation events can be observed in Hi-C contact maps. Chromosomal rearrangement events are shown using karyoplots. Blue density maps represents repeat densities, pink links show syntenic regions. Red links represent inversions.

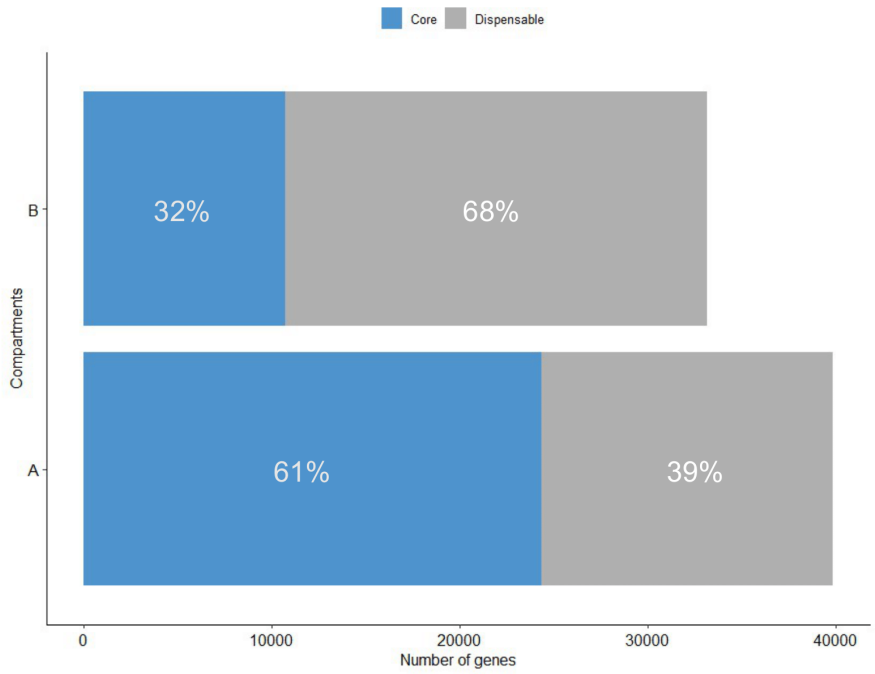
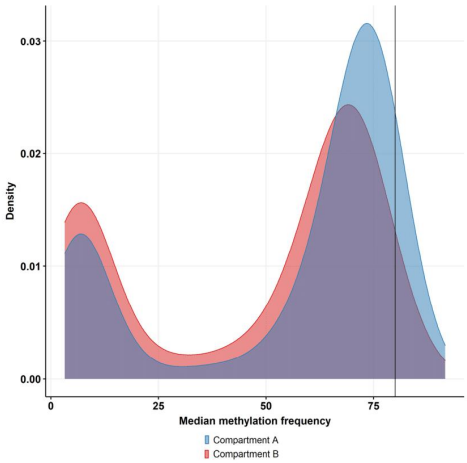
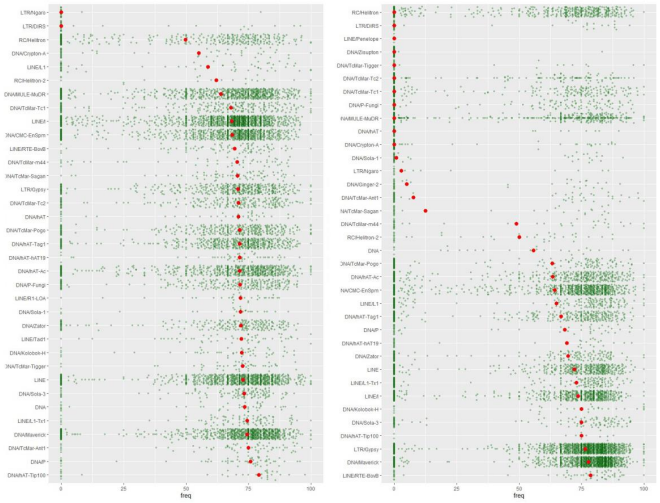

Core
Dispensable
Compartments
B
A
32%
68%
61%
39%
0
10000
20000
30000
40000
Number of genes

Figure S2.12: Distribution of core and dispensable genes in compartments A and B. FastOrtho output was parsed based on compartment locations. The sum of all the shared genes are shown in blue, and the rest of the genes are shown in grey.

A



B



Compartment A Compartment B

Figure S2.13: Methylation analyses of DAOM197198. A) Median methylation frequency density for each compartment. Compartment A is shown in red, and compartment B is shown in blue. B) Methylation frequencies of the most abundant transposable elements in compartment A and B. Green points represent the methylation frequency of each CpG sites located on the repeats. Median of the methylation frequencies on each CpG site is shown in red.

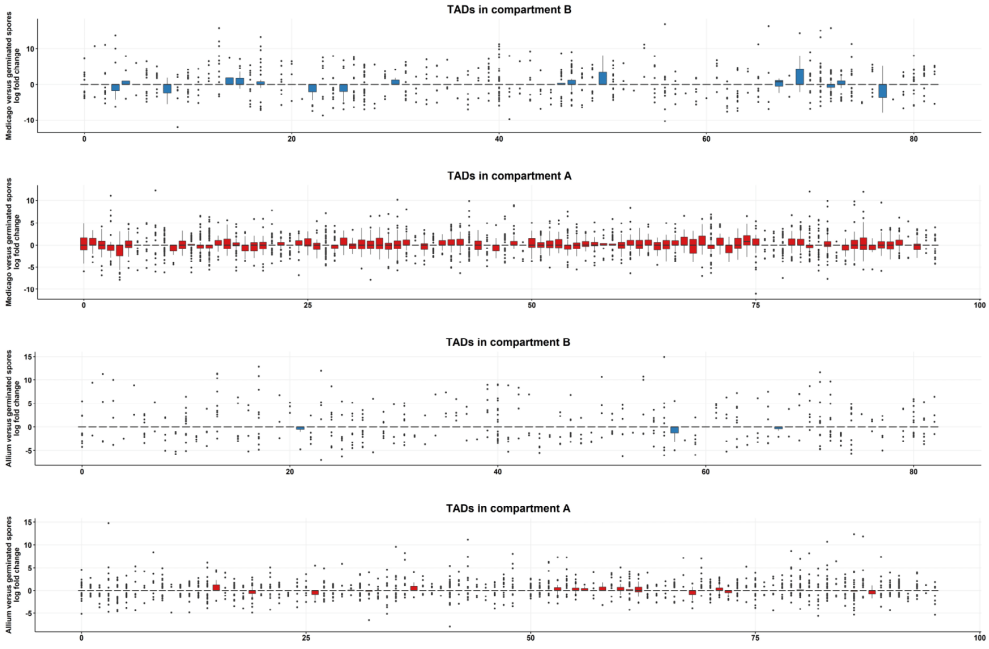

TADs in compartment B
Medicago versus germinated spores
log fold change
10
0
-10
0
20
40
60
80
TADs in compartment A
Medicago versus germinated spores
log fold change
10
5
0
-5
-10
0
25
50
75
100
TADs in compartment B
Allium versus germinated spores
log fold change
15
10
5
0
-5
0
20
40
60
80
TADs in compartment A
Allium versus germinated spores
log fold change
15
10
5
0
-5
0
25
50
75
100

Figure S2.14: Gene expression analysis for the TAD within the A/B compartments. Log fold changes are calculated by comparing extraradical transcription to Medicago and Allium in planta transcription for each compartment.

Table S2.1: Location and number of ribosomal RNA operons within and across strains analyzed in this study.

strain	No of operon copies in genome	Assigned cluster Id	location
DAOM197198	10	cluster1	chr23:3251119-3256917
		cluster2	chr23:3423721-3417923
		cluster3	chr18:1430892-1436711
		cluster4	chr18:1518716-1512898
		cluster5	chr18:1268275-1262457
		cluster6	chr18:1697912-1703739
		cluster7	chr18:1799543-1793726
		cluster8	chr28:2196415-2202234
		cluster9	chr28:2021673-2015854
		cluster10	chr9:825048-819244
C2	9	cluster1	C2_chr_23:3426361-3420560
		cluster2	C2_chr_27:1357134-1351334
		cluster3	C2_chr_18:1599155-1593341
		cluster4	C2_chr_18:2094826-2100641
		cluster5	C2_chr_18:1938983-1933165
		cluster6	C2_chr_18:2197104-2191285
		cluster7	C2_chr_28:1930385-1936203
		cluster8	C2_chr_17:16369-10555
		cluster9	C2_chr_9:4831385-4837185
4401	9	cluster1	4401_chr_6:1203576-1209374
		cluster2	4401_chr_17:104318-98521
		cluster3	4401_chr_18:1278486-1272659
		cluster4	4401_chr_18:1566041-1571869
		cluster5	4401_chr_18:1456630-1462449
		cluster6	4401_chr_18:1658101-1652271
		cluster7	4401_chr_28:2176773-2182584
		cluster8	4401_chr_28:2058040-2052220
		cluster9	4401_chr_9:4306743-4312545
B3	11	cluster1	B3_chr_17:105063-99265
		cluster2	B3_chr_23:3185579-3191376
		cluster3	B3_chr_18:1876955-1882774
		cluster4	B3_chr_18:1973999-1968180
		cluster5	B3_chr_18:1703895-1709712
		cluster6	B3_chr_18:2087711-2093529
		cluster7	B3_chr_18:1511307-1505492
		cluster8	B3_chr_18:1743355-1749207
		cluster9	B3_chr_28:2100828-2095005
		cluster10	B3_chr_28:2264812-2270626

		cluster11	B3_chr_9:868306-862504
A1	10	cluster1	A1_chr_23:2999687-3005484
		cluster2	A1_chr_17:103825-98029
		cluster3	A1_chr_18:1961292-1955474
		cluster4	A1_chr_18:1764224-1758405
		cluster5	A1_chr_18:1369078-1363247
		cluster6	A1_chr_18:1563555-1569386
		cluster7	A1_chr_18:1864737-1870568
		cluster8	A1_chr_28:2358922-2364735
		cluster9	A1_chr_28:2164999-2159179
		cluster10	A1_chr_9:924538-918738

Table S2.2: A/B compartment properties in the model strain DAOM197198.

	A-compartment	**B-compartment**
Size (Mb)	58.1 Mb	73.5 Mb
GC content	28.10%	27.40%
Average genes per 10 Kb	2.55	1.74
Average repeats per 10 Kb	18.21	18.9
Average repeats (excluding simple repeats) per 10 Kb	12.56	14.99
All proteins	12,975	11,334
Secreted proteins	341 (2.6%)	355 (3.1%)*
Apoplastic effector candidates (EffectorP 3.0)	28 (0.2%)	52 (0.5%)*
Cytoplasmic effector candidates (EffectorP 3.0)	164 (1.3%)	187 (1.7%)*

Table S2.3: Comparison of A and B compartment similarity in DAOM197198, C2 and A1 strains.

	A-compartment: DAOM vs. A1	B-compartment: DAOM vs. A1	A-compartment: DAOM vs. C2	B-compartment: DAOM vs. C2
SNPs per Mb	3,847	7,516	8,497	10,636
Indels per Mb	4,182	10,278	10,774	15,389
Average 1-1 alignment length	7,075 bp	3,236 bp	2,811	2,167
Average 1-1 alignment identity	98.90%	97.20%	97%	95.50%
Breakpoints per Mb	797	1,279	1,641	1,709
Relocations per Mb	3	9	5	12
Translocations per Mb	42	87	122	121
Inversions per Mb	2	6	4	8

Author contributions

GY acquired the data. JS and GY performed the main experiments and analyses presented here. ECCH and MM carried some of the analyses. CC produced the fungal material necessary to acquire the data. NC wrote the paper with the continued help of GY and JS. WI supervised the work from ECCH. All authors provided comments on the paper.

Data availability

Reads and fasta files used to generate the analyses are available in Genbank under the BioProject PRJNA748024. Chromosome annotations are available using the doi:10.5281/zenodo.5181509.

Chapter 3: Hi-Fi long reads and Hi-C sequencing resolve two complete and functionally distinct parental haplotypes in arbuscular mycorrhizal fungal heterokaryons

Jana Sperschneider [1*,] Gokalp Yildirir [2*] Mathu Malar C [2,] Pedro Talhinhas [3], Eric CH Chen [4], Essam Sorwar [2], and Nicolas Corradi [2 #]

[1] Black Mountain Science and Innovation Park, CSIRO Agriculture and Food, Canberra, Australia

[2] Department of Biology, University of Ottawa, ON, Ottawa, K1N 6N5, Canada

[3] LEAF-Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal

[4] The University of Tokyo, Department of Integrated Biosciences, Graduate School of Frontier Sciences

* contributed equally

Corresponding Author

Abstract

Arbuscular mycorrhizal fungi (AMF) represent plant symbionts with a multinucleate cytoplasm. Based on fragmented genome and single nucleus data, it was proposed that some strains, referred to as AMF dikaryons, carry thousands of nuclei derived from two parental genotypes in a large syncytium. Here, we set to investigate the origin, diversity and genome biology of this remarkable genetic condition by sequencing four AMF dikaryons using HiFi PacBio and Hi-C sequencing. We find that these strains carry two sets of parental chromosomes. Parental haplotypes are more closely related phylogenetically than to any other sequenced relative, but they differ in content, epigenetics, and relative coverage. No additional haplotypes are assembled, however single nucleus data supports rare events involving somatic recombination events in these strains. This work conclusively supports current models that explain AMF nuclear organization. Findings also reveal the evolutionary origin of co-existing parental haplotypes and their relative genetic contribution to each strain, opening avenues for the environmental applications of AMF dikaryons.

Introduction

Arbuscular mycorrhizal fungi are prominent root symbionts that improve access to nutrients and resistance against pathogens for most land plants (Bonfante & Genre, 2010; Corradi & Bonfante, 2012; Martin *et al.*, 2017). These keystone mutualists increase ecosystem productivity in global terrestrial ecosystems (Lee *et al.*, 2013) and many industries specialize in their production as bio-stimulants for agricultural practices, forestry, and plant nurseries (Vosátka *et al.*). AMF cells are always multinucleate, with some species carrying over 20,000 nuclei within individual

spores (Kokkoris *et al.*, 2020)., and as opposed to other multinucleate fungi, there is no stage in the AMF life cycle where one or two nuclei co-exist in one cell (Kokkoris *et al.*, 2020).

Fossil evidence indicates this plant-fungus symbiosis dates back more than 400 million years (Redecker *et al.*, 2000; Walker *et al.*, 2018; Delaux & Schornack, 2021), but despite this longevity, no sexual stage has yet been formally observed in these fungi. Because of this, AMF were long pigeonholed in a select group of organisms called "ancient asexuals" (Gordo & Charlesworth, 2000; Normark *et al.*, 2003; Schurko *et al.*, 2009), and this long-term absence of sexual reproduction was proposed to be offset by large genomic divergence among their co-existing nuclei (Kuhn *et al.*, 2001; Sanders, 2002; Hijri & Sanders, 2005; Sanders & Croll, 2010).

The view that these organisms have peculiar biology and genetics has, however, been challenged with the discovery of AMF gene orthologues that function in meiosis (Halary *et al.*, 2011; Malar C *et al.*, 2021, 2022), and more recently by evidence that some strains, referred to as AMF dikaryons (or AMF heterokaryons) carry thousands of nuclei derived from two parental strains (Ropars *et al.*, 2016; Chen *et al.*, 2018b,a); a condition analogous to the sexual (heterokaryotic) stage of basidiomycetes (Xu *et al.*, 1996; Anderson & Kohn, 2014).

AMF dikaryons share remarkable ecological and cellular traits with conventional fungal dikaryons, which carry 2 nuclei per cell (Kokkoris *et al.*, 2020). For example, in fungal dikaryons, the co-existence of two genetically different parental nuclei improves fitness and adaptability to environmental change (Springer-Verlag *et al.*, 1988; Clark & Anderson, 2004; Gehrmann *et al.*, 2018), and this ability correlates with change in the relative ratio of their co-existing nuclei (Jinks, 1952). In AMF dikaryons many life-history traits, including spatial hyphal exploration rates, also differ significantly compared to their homokaryotic relatives (Clark & Anderson, 2004; Gehrmann

et al., 2018), and there is also evidence that parental nuclei vary in abundance depending on the strain identity, environmental factors, and plant hosts (Kokkoris *et al.*, 2021; Cornell *et al.*, 2022).

In the dikaryotic basidiomycete *Agaricus bisporus* (Gehrmann *et al.*, 2018), co-existing nuclei express different genes throughout the life cycle of this species, highlighting the different molecular role parental genotypes plays in the development of fungal dikaryons. However, whether similar mechanisms also exist in AMF dikaryons is unknown because information on parental haplotypes are currently unavailable for these strains. As such, the relative genetic, epigenetic, and transcriptional contribution of co-existing haplotypes to the biology of AMF dikaryons cannot be investigated.

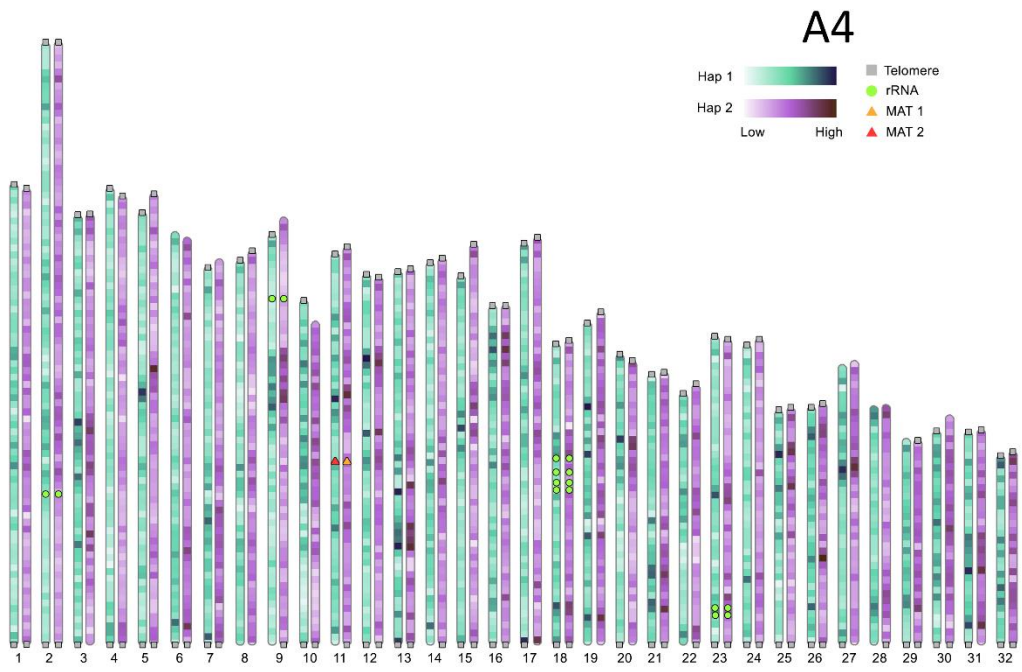
In addition to reveal their genetic content, complete information from these haplotypes would also reveal the evolutionary origin of these strains and indicate if these emerged through sexual or somatic hybridization events (Li *et al.*, 2019). Such data would also help demonstrate whether these strains carry two dominant distinct haplotypes, as proposed by Ropars and colleagues (Ropars *et al.*, 2016) or, a mixture of highly divergent and recombining haplotypes, as recently suggested (Limpens *et al.*, 2022). To address these questions, we sequenced three publicly available AMF dikaryons using High Fidelity (HiFi) PacBio long reads and Hi-C sequencing.

Results

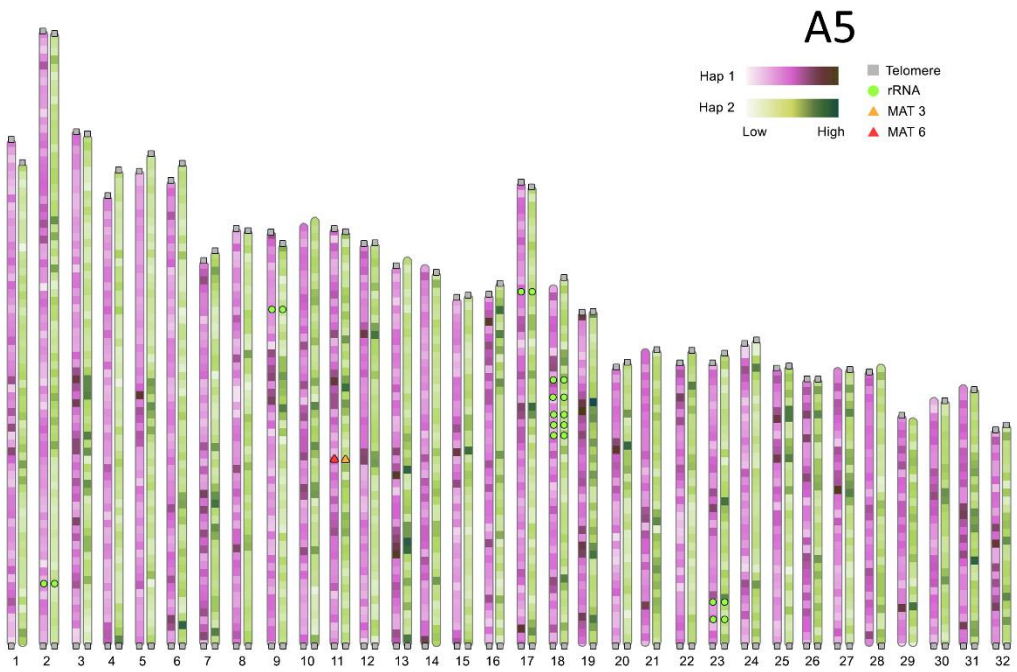
HiFi PacBio long-reads and Hi-C sequencing data was obtained for the publicly available AMF dikaryons A4 (DAOM-664343), A5 (DAOM-664344) and G1 (DAOM-97085). This data was assembled with Hifiasm (Cheng *et al.*, 2021), which in Hi-C mode returned two phased haplotypes. Besides contaminants, no additional haplotype was assembled. Phasing of the

assembled contigs was confirmed with NuclearPhaser and synteny plots made by D-Genies confirmed complete phases (Cabanettes & Klopp, 2018; Duan *et al.*, 2022) **(Supplemental Figure 3.1)**. Each haplotype assembly was then further scaffolded into 32 chromosome-scale scaffolds, guided by Hi-C data (**Figure 3.1a,b,c**). The size of parental haplotypes is consistent with flow cytometry analyses (Ropars *et al.*, 2016), and haplotypes have BUSCO gene repertoire completeness of at least 95.5% (**Table 3.1)**. Overall, the chromosomes of parental haplotypes are highly syntenic with the distant *R. irregularis* homokaryons, and rearrangement events are rare among chromosomes in all strains **(Supplemental Figure 3.2)**. The parental haplotypes of all AMF dikaryons carry the expected number of genes and rRNA operons (**Table 3.1; Supplemental Table 3.1).** Genome coverage analyses of parental haplotypes support recent molecular findings that these differ in relative abundance between AMF dikaryons (**Table 3.1)** (Kokkoris *et al.*, 2021) .

a)


A4
Hap 1
Hap 2
Low
High
Telomere
rRNA
MAT 1
MAT 2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32

b)


A5
Hap 1
Hap 2
Low
High
Telomere
rRNA
MAT 3
MAT 6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32

c)

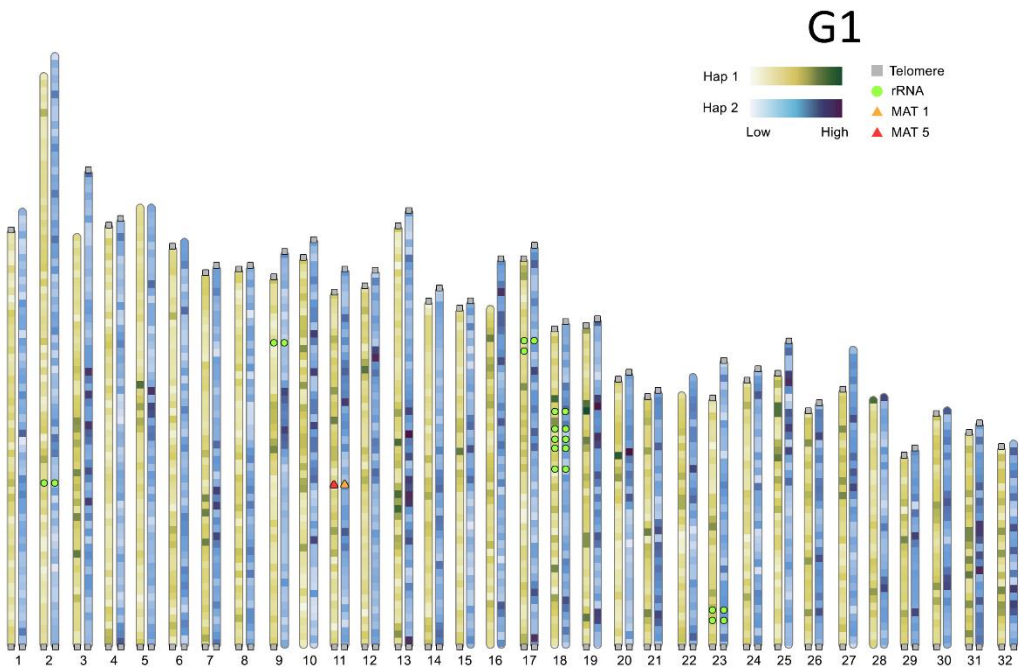

G1
Hap 1
Hap 2
Low
High
Telomere
rRNA
MAT 1
MAT 5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32

Figure 3.1a,b,c: Karyoplots representing the 32 chromosomes belonging to *Rhizophagus irregularis* dikaryotic strains A4, A5 and G1. Both haplotypes that coexist in dikaryotic strains are shown side by side. The color darkness represents gene densities, where lighter colors represent low gene density, and darker colors indicate higher gene density regions. Telomeres, rRNA operons and MAT loci are shown.

Table 3.1: Assembly statistics for the three chromosome-scale, fully phased dikaryons A4, A5 and G1.

	A4		A5		G1	
Statistic	**Hap1**	**Hap2**	**Hap1**	**Hap2**	**Hap1**	**Hap2**
Flow cytometry genome size	140.6 Mb		132.6 Mb		187 Mb	
Assembly size	159.13 Mb	158.70 Mb	147.06 Mb	145.94 Mb	143.49 Mb	145.29 Mb
# of scaffolds	32	32	32	32	32	32
Assembly N/L50	5.5 Mb/13	5.61 Mb/13	4.88 Mb/13	4.93 Mb/13	4.93 Mb/13	5.01 Mb/13
Mating type locus	MAT-2	MAT-1	MAT-6	MAT-3	MAT-5	MAT-1
Average HiFi read coverage	85.8	78.8	111.4	108.4	124.1	76.1
GC content	27.75%	27.76%	27.84%	27.77%	27.54%	27.52%
Complete BUSCOs	96.30%	95.60%	96.00%	95.90%	96.10%	96.00%
# of genes	23078	23095	22583	22191	24143	24515
% of repetitive sequence	47.52%	47.45%	46.63%	46.31%	46.25%	46.00%

The Hi-C maps from each haplotype confirm that the *Rhizophagus irregularis* chromosomes separate into two dominant euchromatin (A) and heterochromatin (B) compartments (Supplemental Figure 3.3). Each compartment significantly differs in average gene and repeat density (Supplemental Figure 3.4a,b) and in methylation (Supplemental Figure 3.4c). These Hi-C maps also reveal differences in chromosomes compartmentalization (and thus in spatial arrangement) between co-existing haplotypes, as previously reported between the genomes of homokaryotic strains (Yildirir *et al.*, 2021) (Supplemental Figure 3.5).

Strains of the model AMF *R. irregulari*s differ significantly in gene content, epigenetics and chromosome size and conformation (Chen *et al.*, 2018b; Yildirir *et al.*, 2021) . These differences also exist between co-existing haplotypes (Figure 3.1, Supplemental Figure 3.6a,b, Supplemental Figure 3.3). For example, the A5 haplotypes differ by more than 2000 genes with motifs predicted to play a role in signaling pathways (e.g. protein kinases), or protein–protein interactions (e.g. the tetratricopeptide repeat Sel1). In A4, protein motifs are found in similar numbers between haplotypes, highlighting inter-strain variability in gene content and plasticity. Haplotype differences in transposable elements and other repeats is also stark. For instance, the LINE/I elements differ by more than 400 in G1, DNA/Tc-MAR elements differ by 2279 in A4, and Pogo DNA transposons by at least 290 in A5 (Table 3.1, Figure 3.1, Supplemental Figure 3.6b).

Each haplotype also carries unique genes and repeats. These genes include secreted proteins and candidate effectors, as well as a few rRNA operons. For instance, G1-Hap1 carries an extra rRNA on Chromosome 17 compared to G1-Hap2. (Supplemental Table 3.1). In addition, A4-Hap 1 carries 4381 genes (including 25 candidate effectors) and 858 TE not found

in A4-Hap 2. Similarly, the A5-Hap 1 carries 3948 genes (1 candidate effector), and 1350 extra TEs, while G1-Hap 1 carries 4654 genes (22 candidate effectors) but 3444 less TEs then G1-Hap 2. In AMF dikaryons, haplotype- specific genes have multiple putative functions and these genes are mostly involved in signal transduction and posttranslational modification functions (Supplemental Table 3.2, Supplemental Figure 3.6a). Notably, differences in gene content and chromosome size among haplotypes are smaller than those observed between homokaryotic relatives; at most, haplotypes differ by an average of 392 genes and 986kb repeats in AMF dikaryons, compared to 2646 genes and 16.39 Mb repeats between the AMF homokaryons studied to date. Similarly, homologous chromosome sizes differ by an average of 101319 bp between co-existing haplotypes, but by 551734 bp in AMF homokaryons.

The chromosomes of parental haplotypes are highly syntenic with each other and other strains, and large inter-haplotype rearrangements are very rare (Supplemental Figure 3.7). At a higher resolution, however, notable differences in gene order can be seen between haplotypes. For example, along the putative mating type locus, only four genes (a choline transporter, the homeodomains HD-1 and HD2, and a phosphoglycerate mutase) are conserved in order between haplotypes (Figure 3.2), while surrounding regions variable between co-existing haplotypes and strains. Similar differences are found in all chromosomes in all strains, highlighting widespread differences in haplotype genome order (Supplemental Figure 3.8).

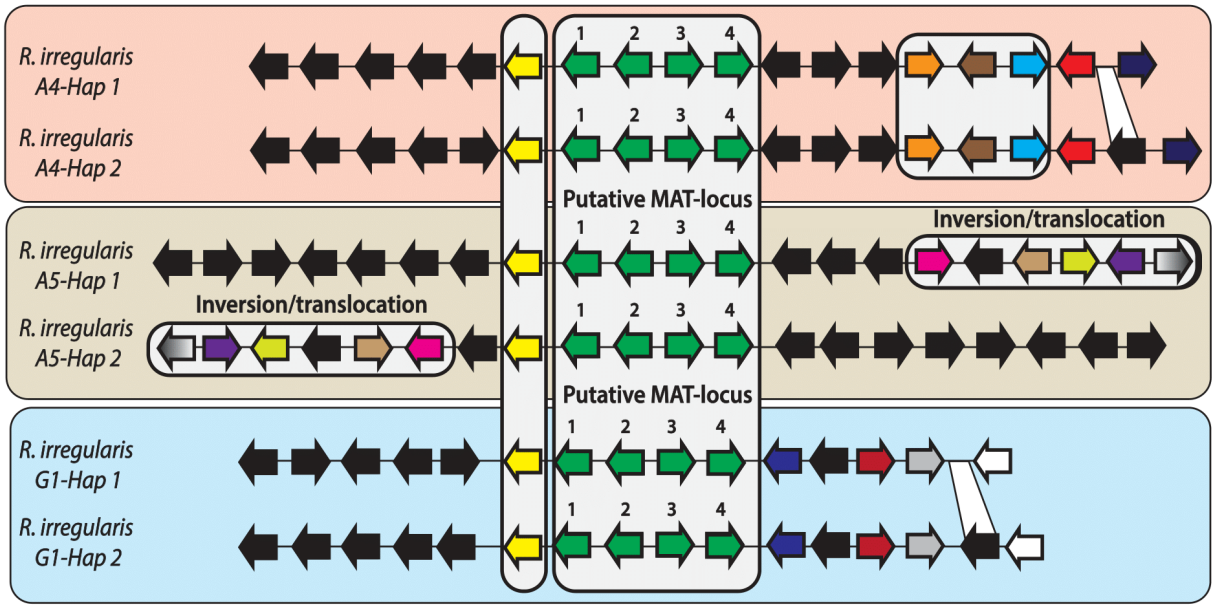

R. irregularis
A4-Hap 1
R. irregularis
A4-Hap 2
1 2 3 4
1 2 3 4
Putative MAT-locus
Inversion/translocation
R. irregularis
A5-Hap 1
R. irregularis
A5-Hap 2
Inversion/translocation
1 2 3 4
1 2 3 4
Putative MAT-locus
R. irregularis
G1-Hap 1
R. irregularis
G1-Hap 2
1 2 3 4
1 2 3 4

Figure 3.2: Genome architecture of putative mating type locus and the surrounding area in haplotypes of *R. irregularis* dikaryotic strains. The arrows represent gene orders and transcriptional directions of genes found in MAT locus. The grey boxes represent syntenic regions. The green arrows represent the putative MAT locus. Black arrows represent unique genes, and same color arrows indicate synteny between two parental haplotypes. Inversion and gene insertion events are marked.

It was hypothesized that AMF dikaryons originate from two compatible homokaryotic strains, which are represented by the co-existing haplotypes we identified. To assess their evolutionary origin, a 105979 bp alignment was generated using their sequences and available genome data from 66 strains catalogued as *R. irregularis* (Savary *et al.*, 2017; Yildirir *et al.*, 2021) (**Figure 3.3**). Maximum likelihood analyses revealed a large phylogenetic diversity in this species (Schüßler *et al.*, 2001; Redecker & Raab, 2017), with at least 6 distinct and highly supported clades. All co-existing haplotypes cluster together and diverge from other sequenced relatives. As such, the parental haplotypes of AMF dikaryons could not be identified with the current dataset. Inspection of aligned sequences did not reveal evidence of recombination between co-existing haplotypes.

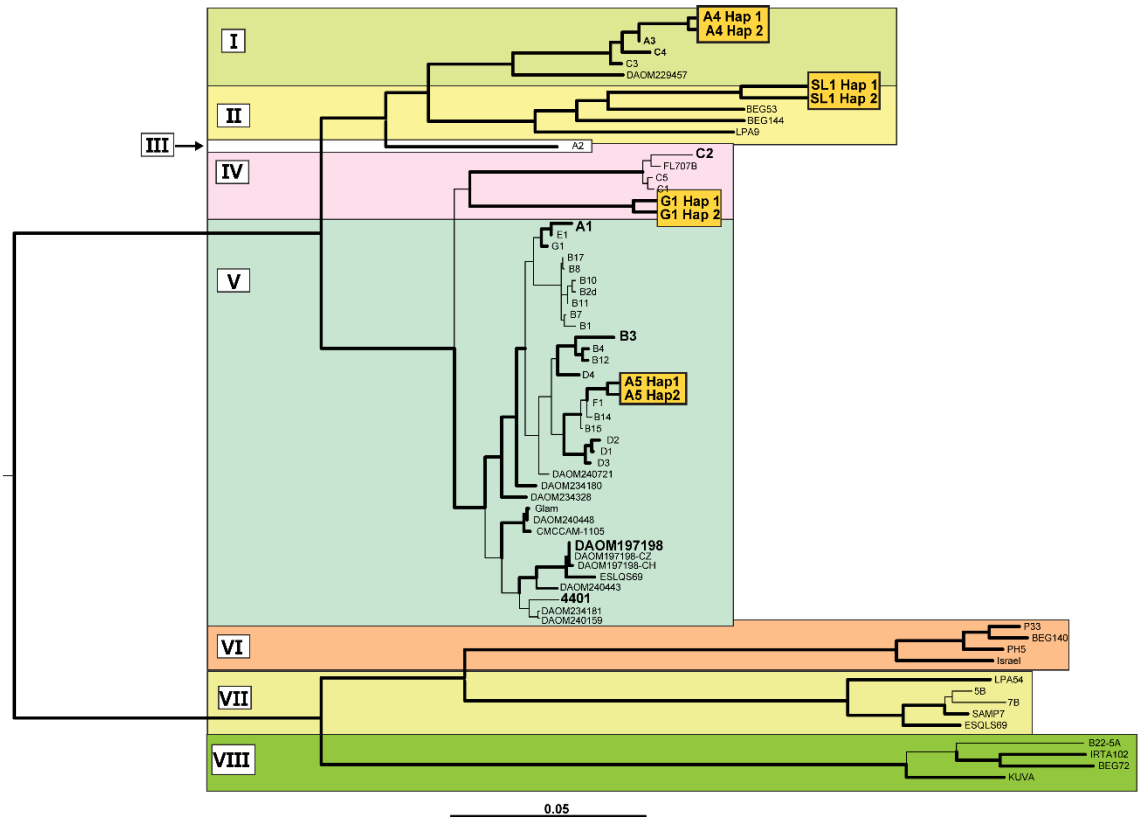

I
II
III
IV
V
VI
VII
VIII
A4 Hap 1
A4 Hap 2
A3
C4
C3
DAOM229457
SL1 Hap 1
SL1 Hap 2
BEG53
BEG144
LPA9
A2
C2
FL707B
C5
C1
G1 Hap 1
G1 Hap 2
A1
E1
G1
B17
B8
B10
B2d
B11
B7
B1
B3
B4
B12
D4
A5 Hap1
A5 Hap2
F1
B14
B15
D2
D1
D3
DAOM240721
DAOM234180
DAOM234328
Glam
DAOM240448
CMCCAM-1105
DAOM197198
DAOM197198-CZ
DAOM197198-CH
ESQLS69
DAOM240443
4401
DAOM234181
DAOM240159
P33
BEG140
PH5
Israel
LPA54
5B
7B
SAMP7
ESQLS69
B22-5A
IRTA102
BEG72
KUVA
0.05

Figure 3.3: Phylogenetic tree constructed from 66 *R. irregularis* strains, achieved by combining genome assemblies and publicly available ddRAD-seq data. Haplotypes from AMF dikaryons are shown in orange squares. All data obtained using chromosome-level is shown in bold. Based on genetic distance, the phylogeny resolves eight distinct clades highlighted in different colors, none of which is shared between AMF dikaryons. Thicker branches indicate bootstrap support > 95. The tree was made using IQTREE algorithm, in GTR-FO mode with 1000 bootstrap replicates.

Based on fragmented and collapsed assemblies it was reported that rare events of recombination occur between nuclei co-existing in the mycelium of AMF dikaryons. These events would allow these strains to generate genetic diversity in these strains(Chen *et al.*, 2018a, 2020; Yildirir *et al.*, 2020) and we aimed to validate these findings using phased datasets. To this end, we mapped available sequences from individual nuclei sequences of A4 and A5 and score their respective SNPs using stringent procedures. This analysis identified multiple cases where individual nuclei shared part of their genotype with those of alternative MAT-locus. These putative cases of inter-nucleus recombinations occur in both strains and can affect over 900kb regions, which can affect up to 197 genes in a single nucleus. However, it is wise to remember that these somatic recombination events only affect 8.5% of available dataset on average and should thus be considered rare (**Figure 3.4**).

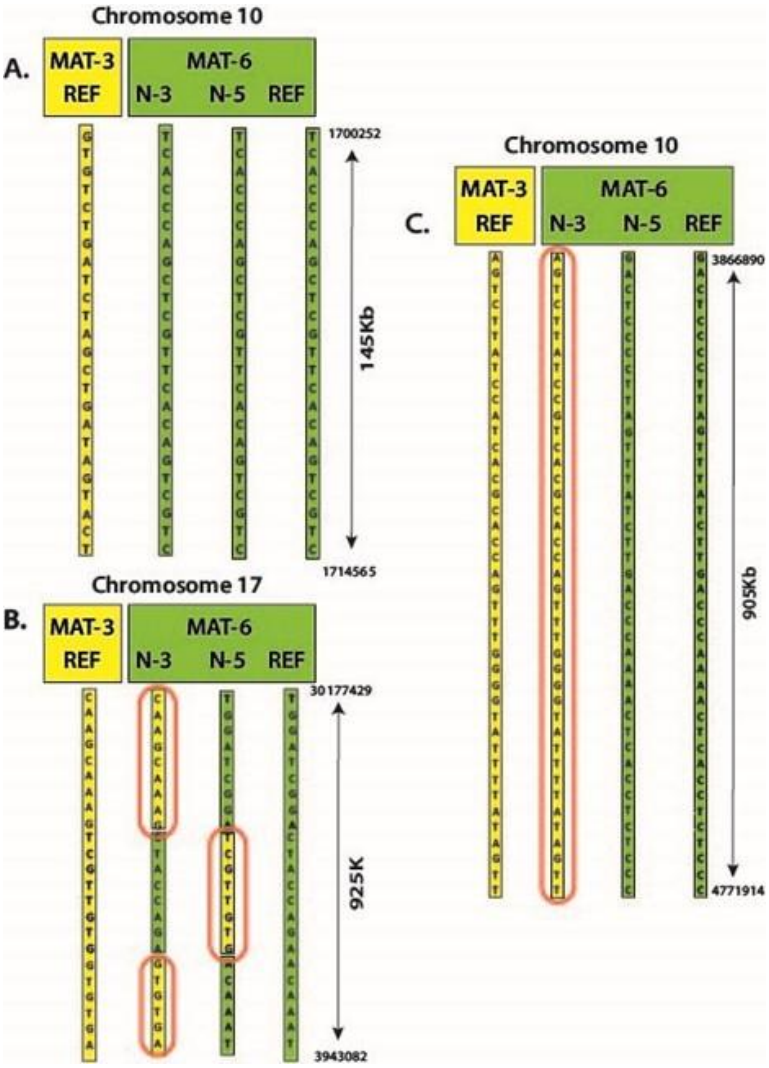

Chromosome 10
A.
MAT-3
REF
MAT-6
N-3
N-5
REF
1700252
145Kb
1714565
Chromosome 17
B.
MAT-3
REF
MAT-6
N-3
N-5
REF
30177429
925K
3943082
Chromosome 10
C.
MAT-3
REF
MAT-6
N-3
N-5
REF
3866890
905Kb
4771914

Figure 3.4: Selected examples of genotypes, recombination and inter-nuclear variability observed in the nuclei 3 and 5 of the dikaryon A5. Recombination events are highlighted in a red circle. These nuclei belong to the MAT-6 nuclei, as confirmed by most of their genome sequence and by PCR with primers based on their MAT-locus(Chen *et al.*, 2018a). Variations along homologous nucleotide positions are highlighted in yellow (MAT-3 genotype) or green (MAT-6 genotype). The SNPs presented in the figure show putative events encompassing hundreds of kilobases and were identified using stringent mapping and scoring procedures.

Discussion

The nuclear organization of AMF has been a contentious scientific topic for decades, with opposing views suggesting that their nuclei are either highly diverging/recombining (i.e. that AMF contain a metagenome) (Sanders, 1999; Kuhn *et al.*, 2001; Hijri & Sanders, 2005) or highly similar in these fungi (Pawlowska & Taylor, 2004, 2005). Recent analyses showed that AMF genome organization follows a homo-heterokaryotic genetic organization (Ropars *et al.*, 2016; Corradi & Brachmann, 2017), however, despite their reconciling views on AMF genetics, these findings were challenged by recent claims that one AMF dikaryon carries highly diverging and recombining genomes (Limpens *et al.*, 2022).

Here, using methods developed to separate haplotypes in complex genomes (e.g. in diploids, polyploids), we conclusively demonstrate that publicly available AMF dikaryons carry thousands of nuclei derived from two parental strains. These findings support the model first published by Ropars and colleagues, which proposes that AMF genetics follow Mendelian patterns and life-stages seen in sexual fungi (Ropars *et al.*, 2016; Corradi & Brachmann, 2017). Our analyses of genetic variability and recombination also confirm that rare somatic recombination events involving co-existing nuclei with distinct genotypes occurs in the mycelium of AMF dikaryons (Chen *et al.*, 2018a, 2020). As such, our data shows that some degree of heterokaryosis is present in these strains. Nevertheless, our findings do not support predictions that long-term absence of sex leads to a substantial accumulation of mutations in AMF nuclei (i.e. Meselson effect)(Welch & Meselson, 2000; Arkhipova & Meselson, 2005) leading to extreme sequence haplotype divergence (Limpens *et al.*, 2022).

Consistent with recent data on AMF homokaryons, co-existing haplotypes carry different genes, repeat and protein domain expansions. This variability in molecular function is

hypothesized to play a role in the adaptability of fungal strains to environmental change (Selmecki *et al.,* 2010; Goodwin *et al.,* 2011; Croll & McDonald, 2012; Stukenbrock, 2013; Plissonneau *et al.,* 2016). The inter-haplotypes gene variability we observed may thus have similar outcomes for AMF dikaryons and is perhaps linked to their ability to grow better across a wider host range (Serghi *et al.*, 2021); a trait that could also translate into higher benefits to host plants.

Hi-C data from AMF dikaryons confirm that haplotype chromosomes carry distinct (A/B) compartments (Yildirir *et al.,* 2021). In *R. irregularis* homokaryons, these compartments have different gene expressions and the repeats are methylated at different rates – i.e. average gene expression and repeat methylation are significantly higher in the A-compartment – and while transcription data is currently unavailable for our AMF dikaryons, the methylation patterns we observed suggest that similar trends are present in these strains as well. Signals from the plant are known to induce changes in chromosome conformation in AMF, leading to the upregulation of genes encoding secreted proteins(Yildirir *et al.*, 2021). As such, it is intriguing to speculate that the differences in chromosome epigenetics we identified among haplotypes (and strains) have a significant role in the establishment and maintenance of the symbiosis in AMF dikaryons.

Although co-existing haplotypes do differ genetically, the co-existing haplotypes are genetically more similar to each other than to any sequenced homokaryotic relatives. This is supported by phylogenetic analyses, which indicate that the AMF dikaryons we studied did not emerge through cross species hybridizations, but via nuclear exchanges between very closely related strains. Future sequencing efforts should now focus on increasing genomic sampling within each of the phylogenetic clades we identified, as well as across natural ecosystems and particularly within populations. This strategy is needed to improve our understanding of AMF biodiversity and provide better species definitions for this group. This approach is also needed to

uncover new AMF dikaryons and reveal their parental homokaryotic strains, and thus ultimately open venues for AMF strain improvement using conventional crossing procedures.

In conclusion, AMF dikaryons carry two sets parental homologous chromosomes that differ in content, architecture, and epigenetics. These strains also carry additional nuclear diversity because of rare somatic recombination and mutations but are otherwise closely related phylogenetically. Although known AMF dikaryons emerge from genetically related strains, no evidence of genetic recombination was found in our dataset. As such, our work suggest that AMF dikaryons originate through somatic hybridization, leading to a cellular situation seen in basidiomycetes where distinct haplotypes co-exist in the same cell over decades without undergoing meiotic recombination (Xu *et al.*, 1996; Li *et al.*, 2019). It will now be important to determine how the genetic differences we detected among haplotypes are transcribed and how they affect the biology of AMF, and ultimately how they involve in the mycorrhizal symbiosis. To this end, future studies should investigate haplotype-specific gene expression using RNA-seq across several host plants, including in-vivo studies, and under different environmental conditions (temperature, pH, salinity). This will reveal which genes are involved in the intricate genetic interactions that occur among haplotypes, and between the partners of the AM symbiosis.

Materials and Methods

Culturing, DNA extraction and ONT/Hi-C sequencing

Rhizophagus irregularis strains A4, A5, and G1 were cultured using either *Daucus carota* or *Cichorium intybus* root organ cultures (ROCs) and as the host, as previously described (Corradi *et al.*, 2004b). The strains are propagated in two-compartment ROCs, allowing us to produce mycelium and spores without obvious contaminants. High-molecular-weight DNA was extracted for all strains using a protocol proposed by Schwessinger & McDonald (Schwessinger, 2016), and was subjected to high fidelity (HiFi) PacBio sequencing at Discovery Life Science (Huntsville, AL). PacBio sequencing produced an average 24,647,320,191 base pairs (bp) of HiFi reads, with a read quality of Q33 and median read length of 13,138.

To obtain high-quality Hi-C data, approximately 200 mg of fresh mycelium from each strain was crosslinked and shipped to Phase Genomics (Seattle, WA, USA) to obtain strain-specific Hi-C data. Hi-C data were produced using the Proximo Hi-C Kit (Microbe) at Phase Genomics with DPNII restriction enzymes. Before generating high-throughput Hi-C Illumina data, the quality of these libraries was assessed by mapping low-coverage paired-end data onto available *R. irregularis* assemblies. A library is deemed to be of high quality if the fraction of high-quality paired-end reads mapping respectively > 10 kb apart within and across publicly available contigs exceeds 1.5% and 5.0%. In all cases, the values obtained for our samples far exceeded the QC limits set by Phase Genomics.

Chromosome assembly and annotation

The HiFi reads of the strains A4, A5 and G1 were assembled with hifiasm 0.16.1 in Hi-C integration mode and with default parameters (Cheng *et al.*, 2021). For G1 --hom-cov was set to 181. The resulting contigs were scanned for possible contaminations and mitochondrial sequences by running a Diamond blastx (diamond v0.9.14.115, Database format version = 1) and flagged contigs were removed from the assembly (Buchfink *et al.*, 2014, 2021). Phasing of the assembled haplotypes was confirmed with the NuclearPhaser pipeline(Duan *et al.*, 2022). For scaffolding, the Hi-C reads were first mapped to each haplotype using BWA-MEM 0.7.17(Li & Durbin, 2009). Alignments were then processed with the Arima Genomics pipeline (Arima Genomics, 2019) Scaffolding was performed using SALSA 2.2 (Ghurye *et al.*, 2017). Hi-C contact maps were produced with HiC-Pro 2.11.1(Servant *et al.*, 2015) (MAPQ=10) and visually examined for correct scaffolding.

De novo repeats were predicted with RepeatModeler 2.0.0 and the option -LTRStruct (Flynn *et al.*, 2020). These were merged with the RepeatMasker repeat library and RepeatMasker 4.1.0 was run with this combined repeat database (http://www.repeatmasker.org).

Gene annotation was achieved by using Funannotate (v.1.7.4) (https://zenodo.org/record/2604804#.YPm336iSnIU) on repeat-masked genome assemblies. Genome annotations were performed using RNA-sequencing (RNA-seq) reads and protein models available for strains A4 and A5 (Grigoriev *et al.*, 2012; Chen *et al.*, 2018a), and using RNA-seq for strain G1. The quality of annotations was evaluated using the Benchmarking Universal Single-Copy Orthologs (Busco, v.5.2.2, database Obd_10) (Simão *et al.*, 2015).

Conserved protein domains were predicted using Pfam v.27 (El-Gebali *et al.*, 2019). SignalP 4.1 (-t euk -u 0.34 -U 0.34)(Petersen *et al.*, 2011) and Tmhmm 2.0 (Krogh *et al.*, 2001) were

used to predict secreted proteins. A protein was called secreted if it was predicted to have a signal peptide and but no transmembrane domains. Effector candidates were predicted with EffectorP 3.0 (Sperschneider & Dodds, 2021). De novo repeats were predicted with RepeatModeler 2.0.0 and the option -LTRStruct (Flynn *et al.*, 2020). These were merged with the RepeatMasker repeat library, and RepeatMasker 4.1.0 was run with this combined repeat database(Smit *et al.*, 2013). TE locations were extracted from the RepeatMasker output file using an R script, and simple repeats, unknown and low-complexity repeats, satellites, tRNA, snRNA and rRNAs were filtered out from the output file. PfamScan(Hancock & Bishop, 2004; Li *et al.*, 2015) was used with default parameters to identify the protein domain annotations in all strains. To create the Pfam heatmaps, the domain numbers were counted for each *R. irregularis* strain or A/B compartment, and a t test conversion was made to highlight the protein domain abundances for each category.

We used mash dist (v.2.2.0) (Ondov *et al.*, 2016) for k-mer distance calculations and dnadiff from the MUMmer(Kurtz *et al.*, 2004; Marçais *et al.*, 2018) suite for structural variation calculations between strains and compartments. Orthology analyses were made by FastOrtho(Wattam *et al.*, 2014) using the following parameters: 50% identity and 50% coverage using protein sequences of all five assemblies. All karyoplots were produced using KaryoploteR(Gel & Serra, 2017). All genome data and reads newly obtained are available in GenBank under the BioProject XXXX.

Identification of chromosome compartments and topologically associated domains

We called A/B compartments and topologically associated domains (TADs) with Hicexplorer 3.6(Ramírez *et al.*, 2018) using the commands hicPCA and hicFindTADs, respectively. Inter- and intrachromosomal Hi-C contact maps were produced with HiC-Pro 2.11.1(Servant *et al.*, 2015) (MAPQ = 10) and Hicexplorer 3.6(Winter *et al.*, 2018; Ramírez *et al.*, 2018; Wolff *et al.*, 2020) and these Hi-C contact maps were manually inspected to assign chromosomal regions to A/B compartments. Specifically, the regions along each chromosomes carrying the same PCA1 values – that is, positive or negative – were assigned to the same compartment. Following this assignment, compartments were manually inspected to investigate their proximity with other chromosomes. Those in close physical proximity with the smallest chromosomes were assigned to compartment A, while those carrying the counterpart PCA1 signal were assigned to compartment B. This process was manually repeated for all 32 chromosomes individually to separate the chromosomal regions into separate A/B compartments. Bedtools(Quinlan & Hall, 2010) was used to assess overlap between genomic features such as genes/repeats and compartments.

Methylation and gene expression analyses

Megalodon (v.2.2.9) was used for methylation calling on on ONT reads using the high-accuracy parameters (dna_r9.4.1_450bps_modbases_dam-dcm-cpg_hac.cfg) configuration file obtained from all strains. After methylation calling was completed for all assemblies, CG positions and their methylation frequencies were extracted from the 5mC bedfile output. For RNA-seq analyses, Salmon v.1.3.0(Patro *et al.*, 2017) was used to align clean RNA-seq reads to the transcripts and to estimate transcript abundances in each sample (salmon index –

keepDuplicates and salmon quant –validateMappings). We used Tximport and DESeq2(Love *et al.*, 2014) to assess gene differential expression (Padj < 0.1).

Phylogenetic analysis

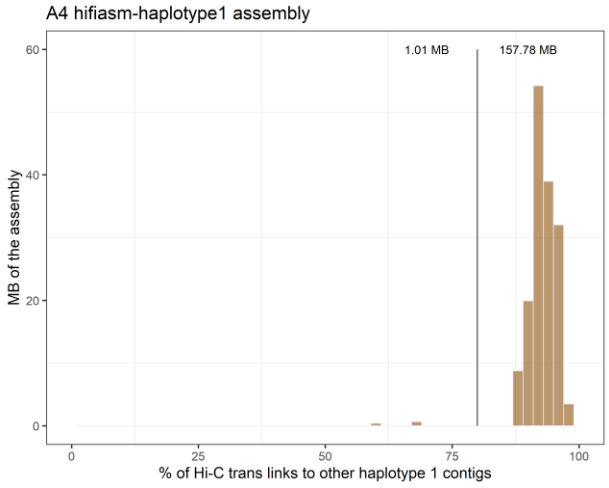
Available ddRAD-seq data sets were included to perform the phylogenetic analysis(Wyss *et al.*, 2016; Savary *et al.*, 2017). Overall, ddRAD-seq data from 60 *R. irregularis* strains with three replicates were used to construct the phylogenetic tree. After the read quality was assessed using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), Illumina adapters were trimmed using TagCleaner(Schmieder *et al.*, 2010), reads containing uncalled bases (N) and shorter reads (<50bp) were removed. The demultiplexing of the dataset was achieved with the *process_radtags* from the Stacks pipeline v2.60(Catchen *et al.*, 2013).

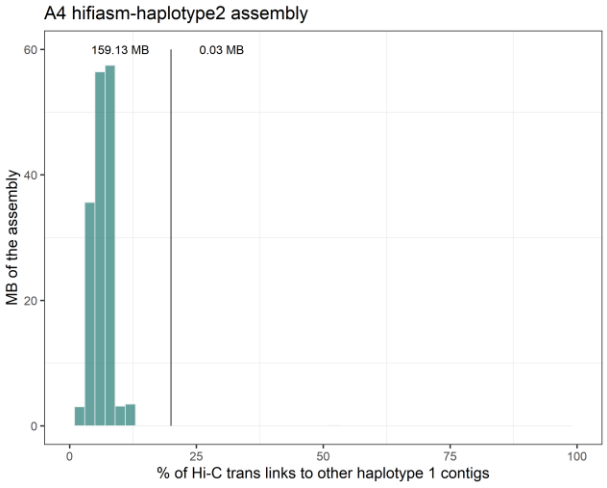
To construct the phylogenetic tree, six haplotype genome assemblies constructed in this study, five homokaryotic assemblies published by Yildirir et al, (Yildirir *et al.*, 2021) and the ddRAD-seq reads that were processed as mentioned above were included in Phame pipeline(Shakya *et al.*, 2020) using these parameters: data = 4, reads = 2, code = 0, cutoff = 0.0001, reference = 1, reffile = DAOM-197198 genome assembly. Using the 105979 bp alignment file, IQTREE algorithm was used with GTR+FO mode with 1000 bootstrap replications to construct the phylogenetic tree(Nguyen *et al.*, 2015). The phylogenetic tree was then further edited by FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

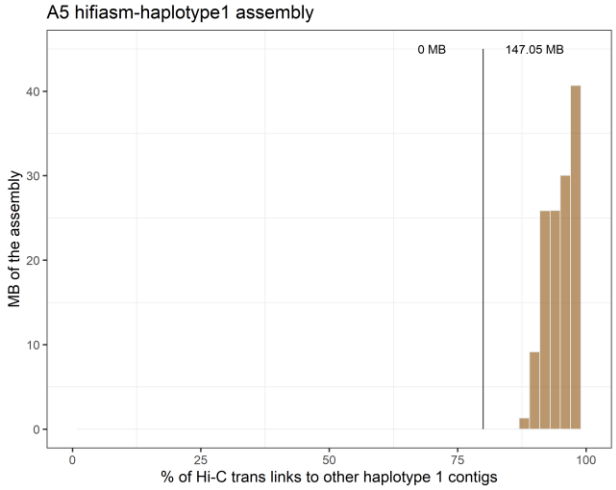
Nuclear DNA content estimation using flow cytometry

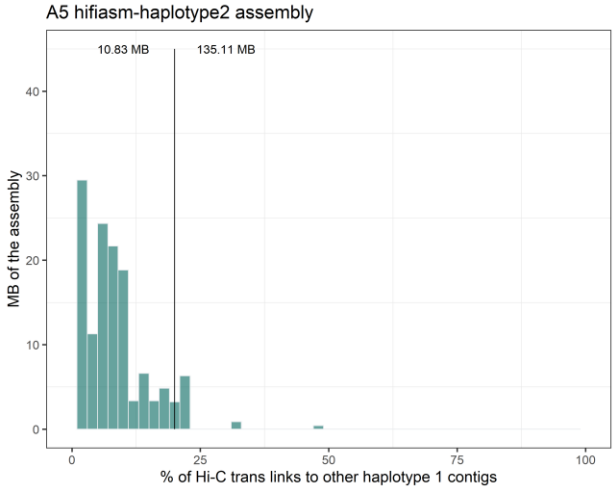
Mycelium of G1 and SL1 (DAOM-240409) strains was recovered from Phytagel medium used in ROC using a needle after incubating slants of gel on citrate buffer (3 mM sodium citrate and 0.5 mM citric acid) for 1 h under gentle agitation. Mycelium of each strain (alone or along a DNA standard) was chopped with a razor blade in the presence of 1 ml of nuclei extraction buffer(Loureiro *et al.*, 2007) following the protocol for filamentous fungi optimized by Pires et al. (Pires *et al.*, 2016). DNA standards used were the fungus *Cenoccocum geophilum* (1C = 0.208 pg DNA)(Talhinhas *et al.*, 2017) and the plant *Rhamnus alaternus* (2C = 0.680 pg DNA) (Carvalho *et al.*, 2018). The nuclear suspension obtained was then separated from plant debris using a 30 μm nylon filter. After filtration, 50 μg/mL of propidium iodide (PI; Sigma-Aldrich) were added to stain DNA and 50 μg/mL of RNase (Sigma-Aldrich) ware added to prevent staining of double stranded RNA. The samples were maintaining at room temperature and analyzed using a CyFlow Space flow cytometer (Sysmex, Germany) equipped with a 30 mW green solid-state laser emitting at 532 nm for optimal PI excitation. Flow cytometer quality assessment, data acquisition parameters (including gating strategies to improve the quality of the histogram) and data analysis were performed as described by Pires et al. (2016)(Pires *et al.*, 2016). At least five replicates were analyzed for each species. The holoploid genome size in pg (1C) (Greilhuber *et al.*, 2005) was calculated using the fluorescence of the sample and the standard, as described by Talhinhas et al. (Talhinhas *et al.*, 2017). The genome size in Mb was also calculated for each strain using the factor 1 pg = 978 Mb (Doležel & Bartoš, 2005).

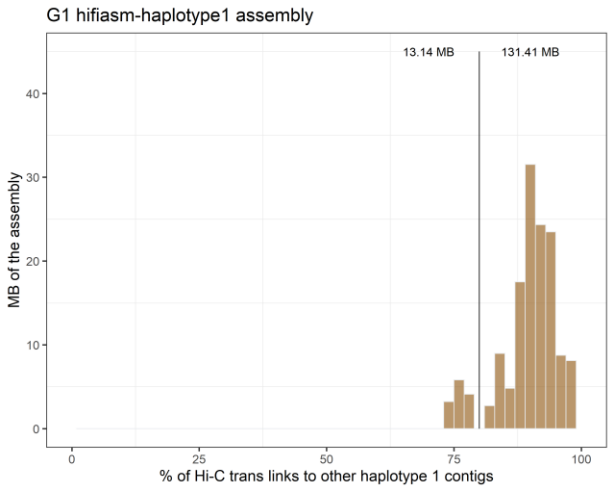
Supporting Figures











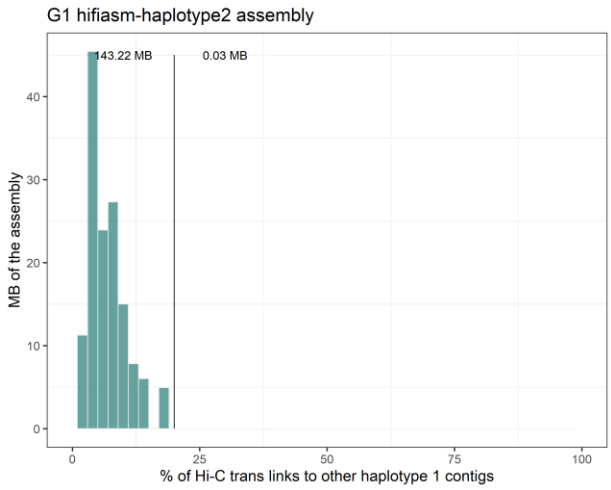


Figure S3.1: The Hifiasm assemblies with Hi-C integration for A4, A5 and G1 are fully phased and exhibit a stron dikaryotic phasing signal.

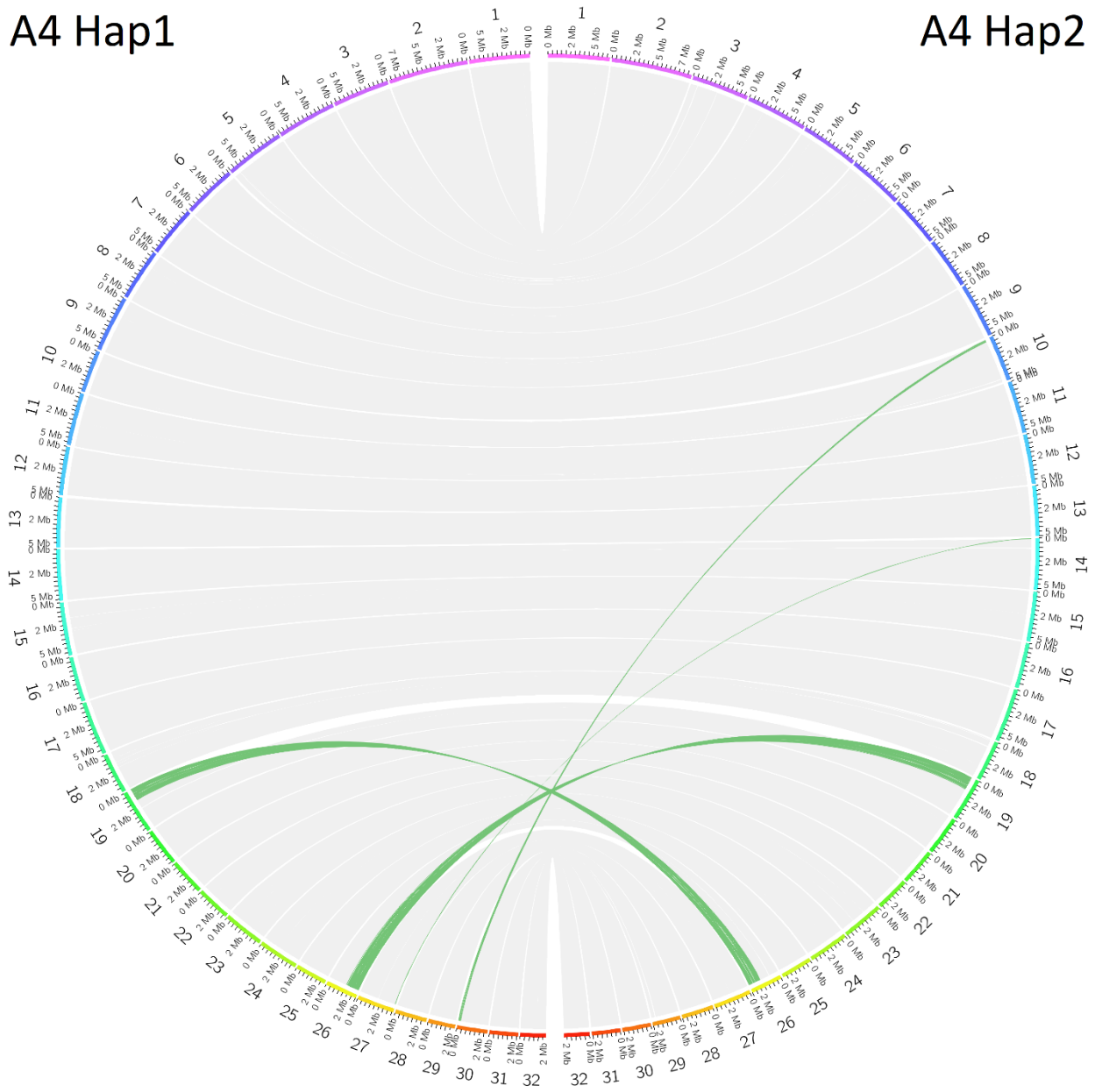
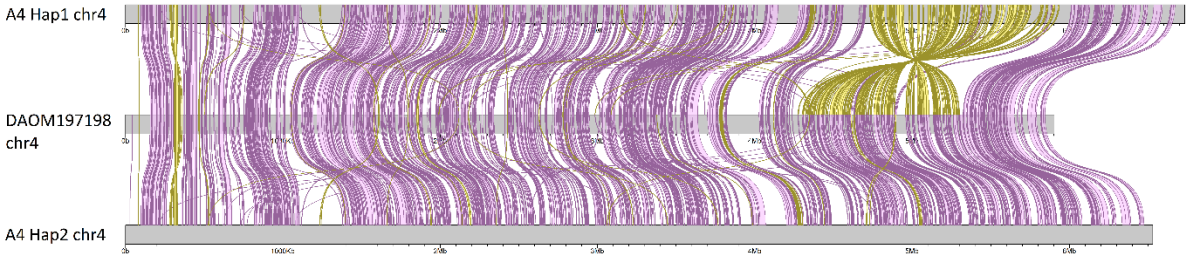
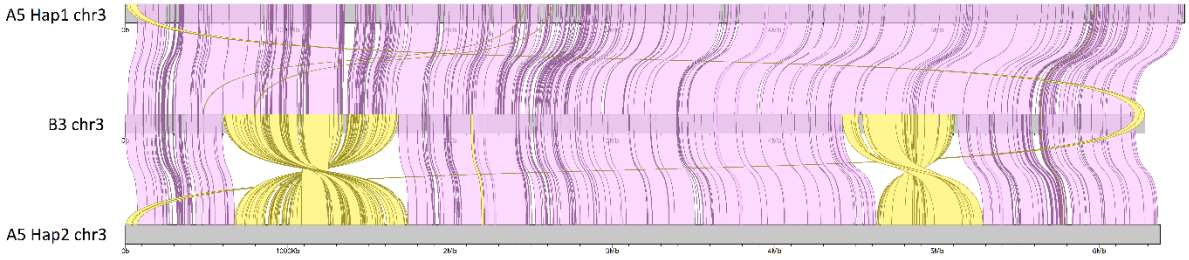

A4 Hap1
A4 Hap2

Figure S3.2: A4 Haplotype 1 and A4 Haplotype 2 assemblies share large synteny blocks in between and rearrangement events are rare. Grey blocks indicate synteny between homologous chromosomes and green blocks show rearrangement events between different chromosomes.


A4 Hap1 chr4
DAOM197198 chr4
A4 Hap2 chr4


A5 Hap1 chr3
B3 chr3
A5 Hap2 chr3

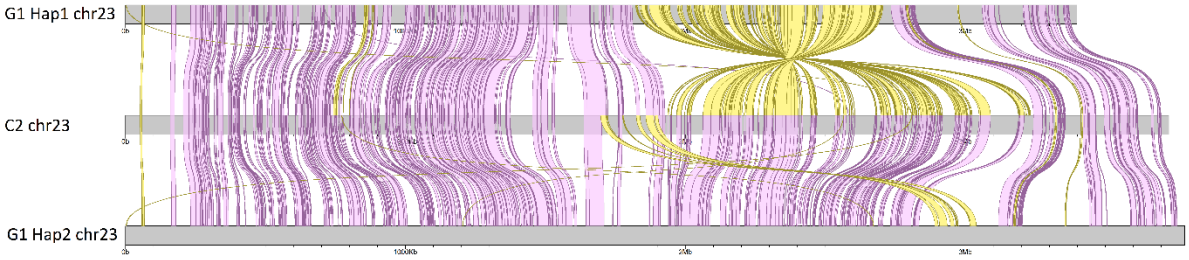
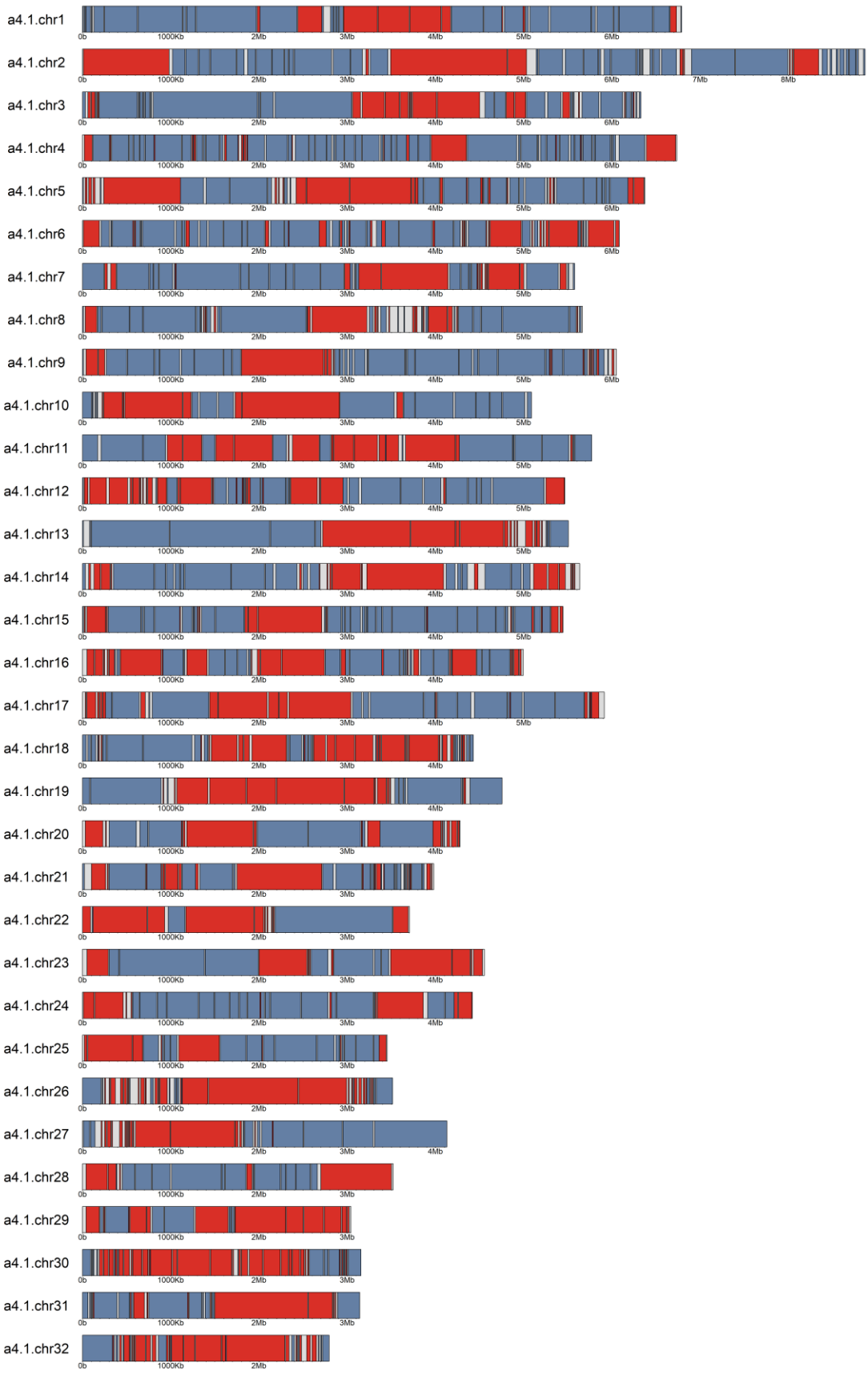
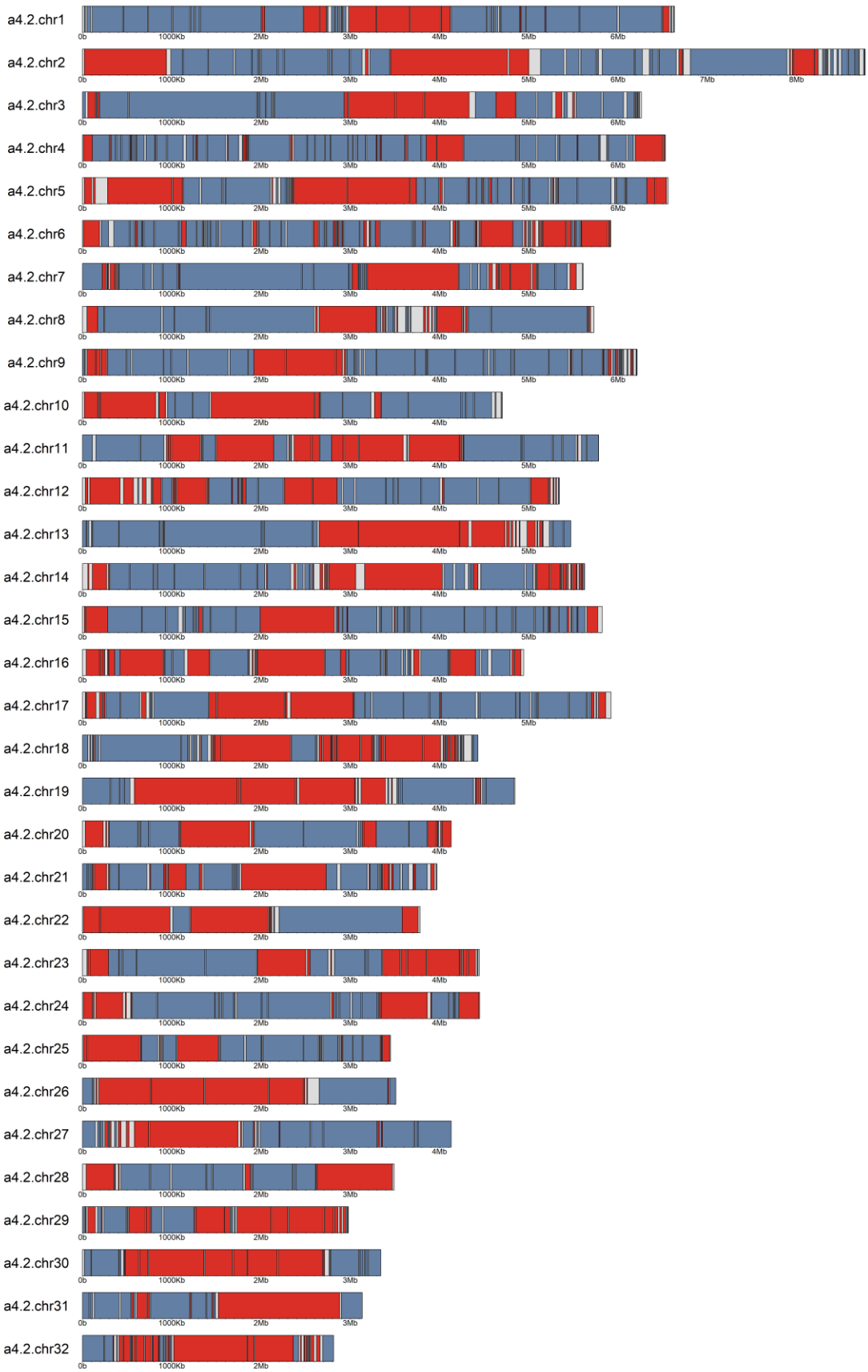
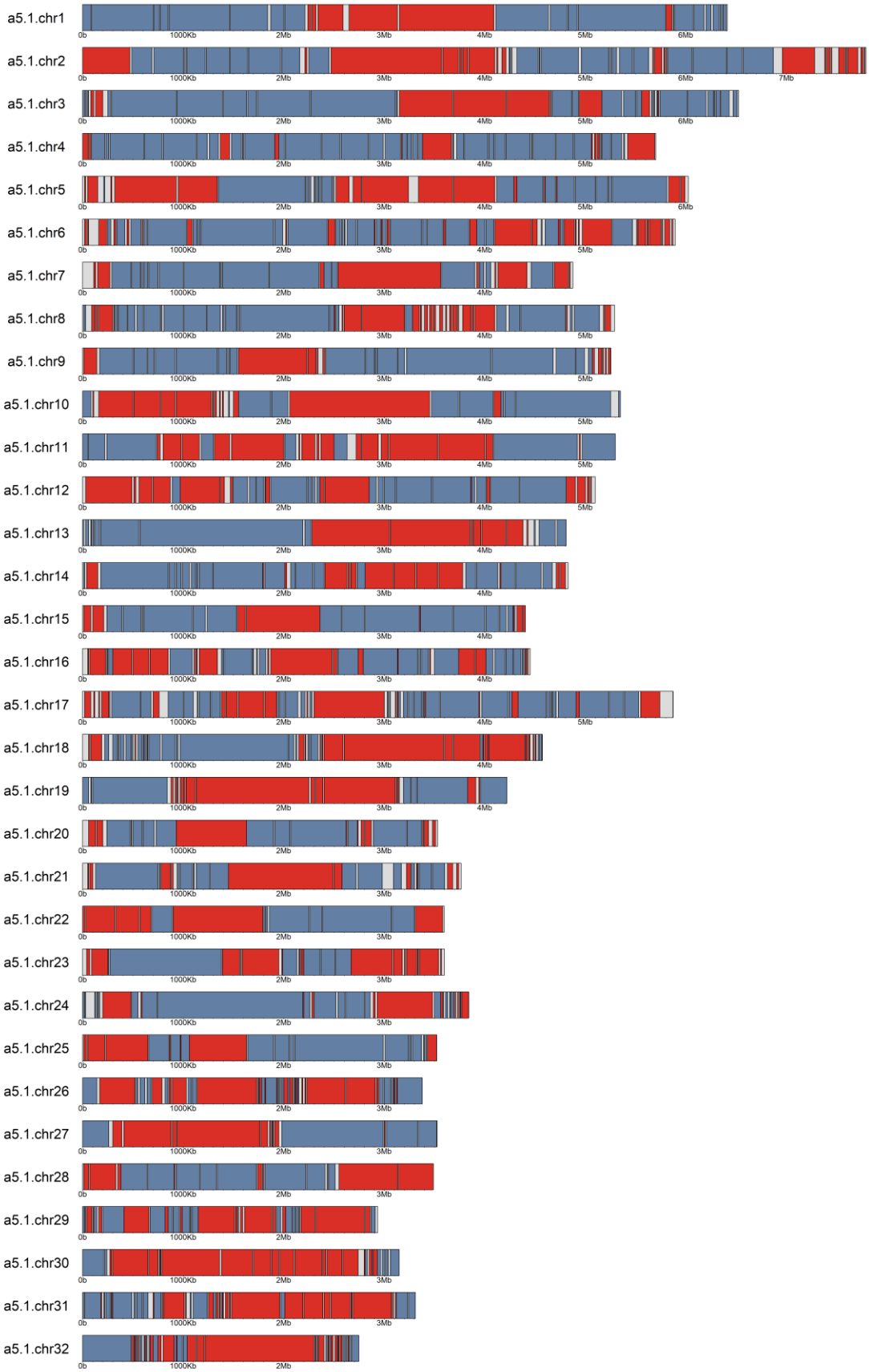

G1 Hap1 chr23
C2 chr23
G1 Hap2 chr23

Figure S3.2: Rare rearrangement events between homologous chromosomes are observed in parental haplotypes when compared to other distant homokaryotic strains. The synteny blocks are shown in pink, inversion events are shown in yellow.


a4.1.chr1
a4.1.chr2
a4.1.chr3
a4.1.chr4
a4.1.chr5
a4.1.chr6
a4.1.chr7
a4.1.chr8
a4.1.chr9
a4.1.chr10
a4.1.chr11
a4.1.chr12
a4.1.chr13
a4.1.chr14
a4.1.chr15
a4.1.chr16
a4.1.chr17
a4.1.chr18
a4.1.chr19
a4.1.chr20
a4.1.chr21
a4.1.chr22
a4.1.chr23
a4.1.chr24
a4.1.chr25
a4.1.chr26
a4.1.chr27
a4.1.chr28
a4.1.chr29
a4.1.chr30
a4.1.chr31
a4.1.chr32
0b
1000Kb
2Mb
3Mb
4Mb
5Mb
6Mb
7Mb
8Mb


a4.2.chr1
a4.2.chr2
a4.2.chr3
a4.2.chr4
a4.2.chr5
a4.2.chr6
a4.2.chr7
a4.2.chr8
a4.2.chr9
a4.2.chr10
a4.2.chr11
a4.2.chr12
a4.2.chr13
a4.2.chr14
a4.2.chr15
a4.2.chr16
a4.2.chr17
a4.2.chr18
a4.2.chr19
a4.2.chr20
a4.2.chr21
a4.2.chr22
a4.2.chr23
a4.2.chr24
a4.2.chr25
a4.2.chr26
a4.2.chr27
a4.2.chr28
a4.2.chr29
a4.2.chr30
a4.2.chr31
a4.2.chr32


a5.1.chr1
a5.1.chr2
a5.1.chr3
a5.1.chr4
a5.1.chr5
a5.1.chr6
a5.1.chr7
a5.1.chr8
a5.1.chr9
a5.1.chr10
a5.1.chr11
a5.1.chr12
a5.1.chr13
a5.1.chr14
a5.1.chr15
a5.1.chr16
a5.1.chr17
a5.1.chr18
a5.1.chr19
a5.1.chr20
a5.1.chr21
a5.1.chr22
a5.1.chr23
a5.1.chr24
a5.1.chr25
a5.1.chr26
a5.1.chr27
a5.1.chr28
a5.1.chr29
a5.1.chr30
a5.1.chr31
a5.1.chr32
0b
1000Kb
2Mb
3Mb
4Mb
5Mb
6Mb
7Mb

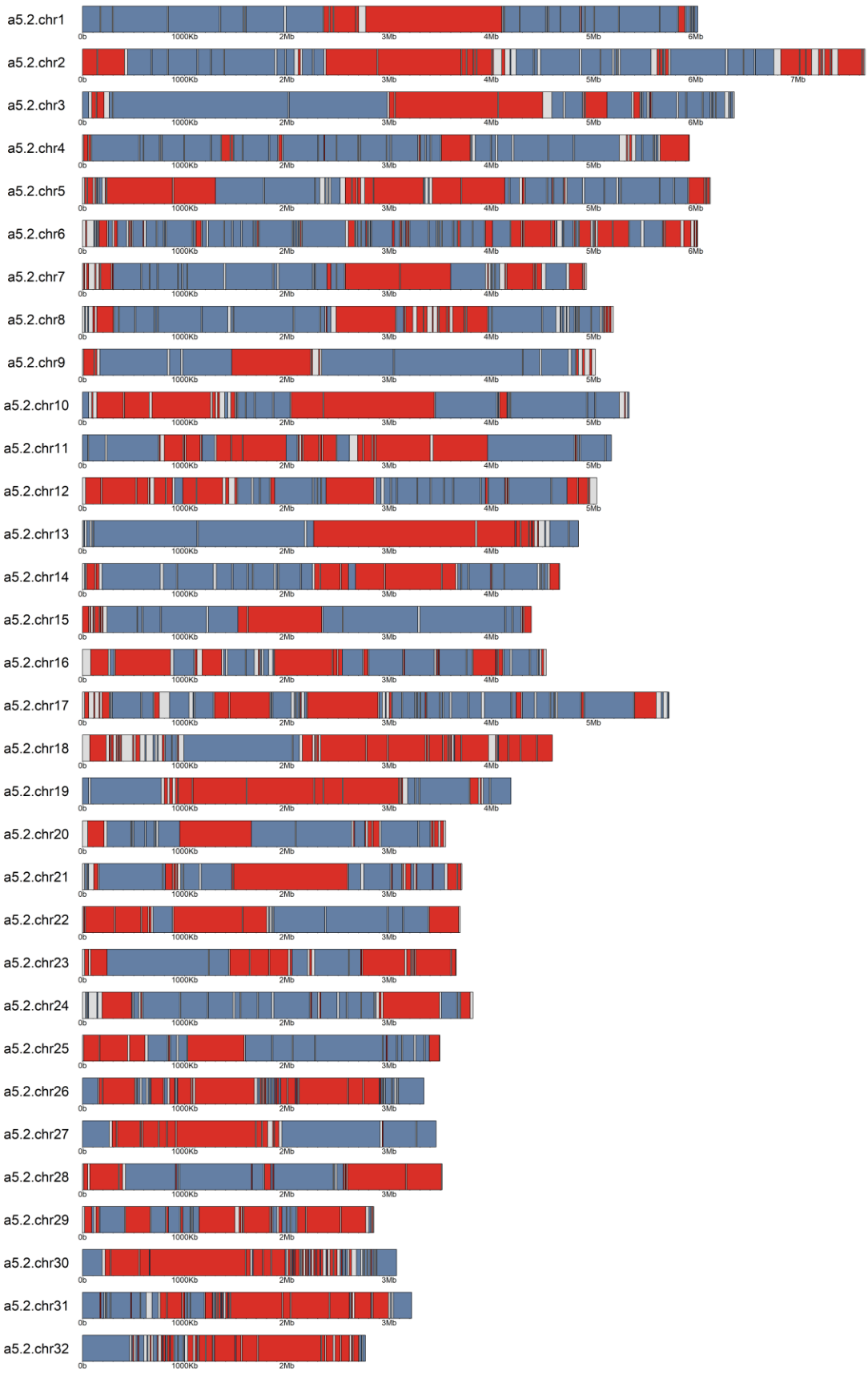
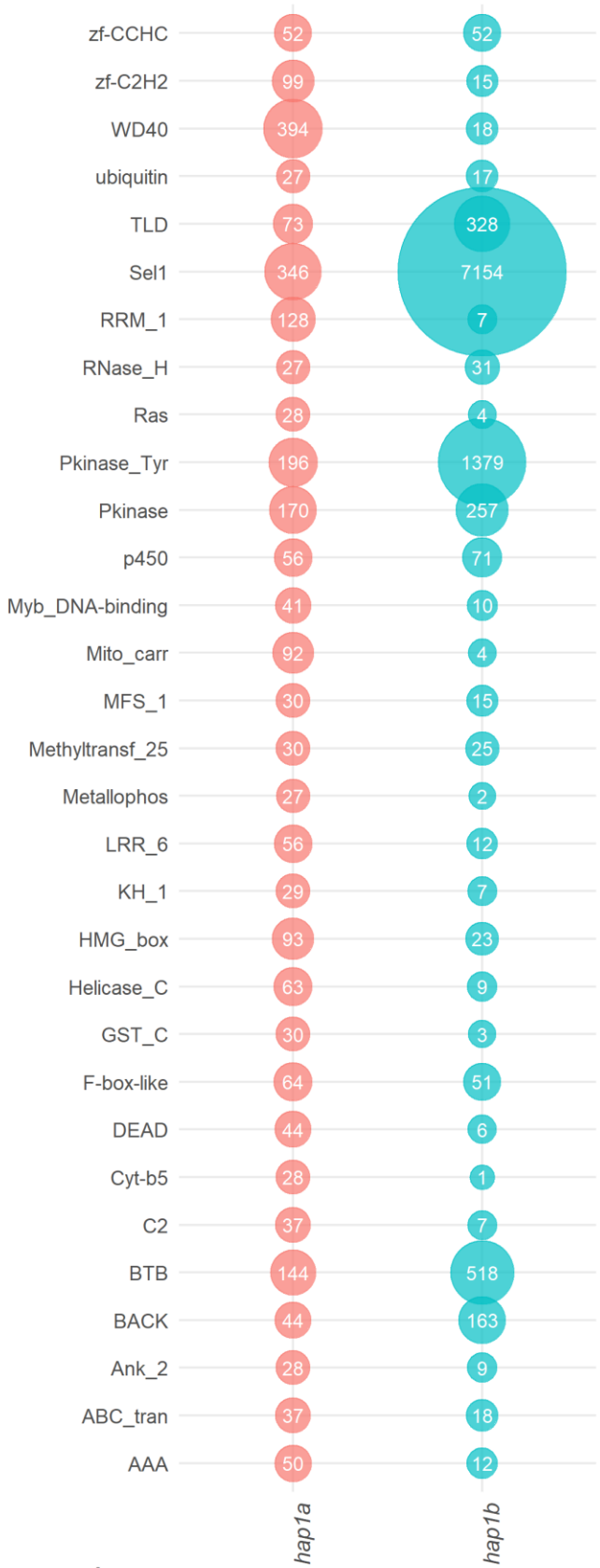
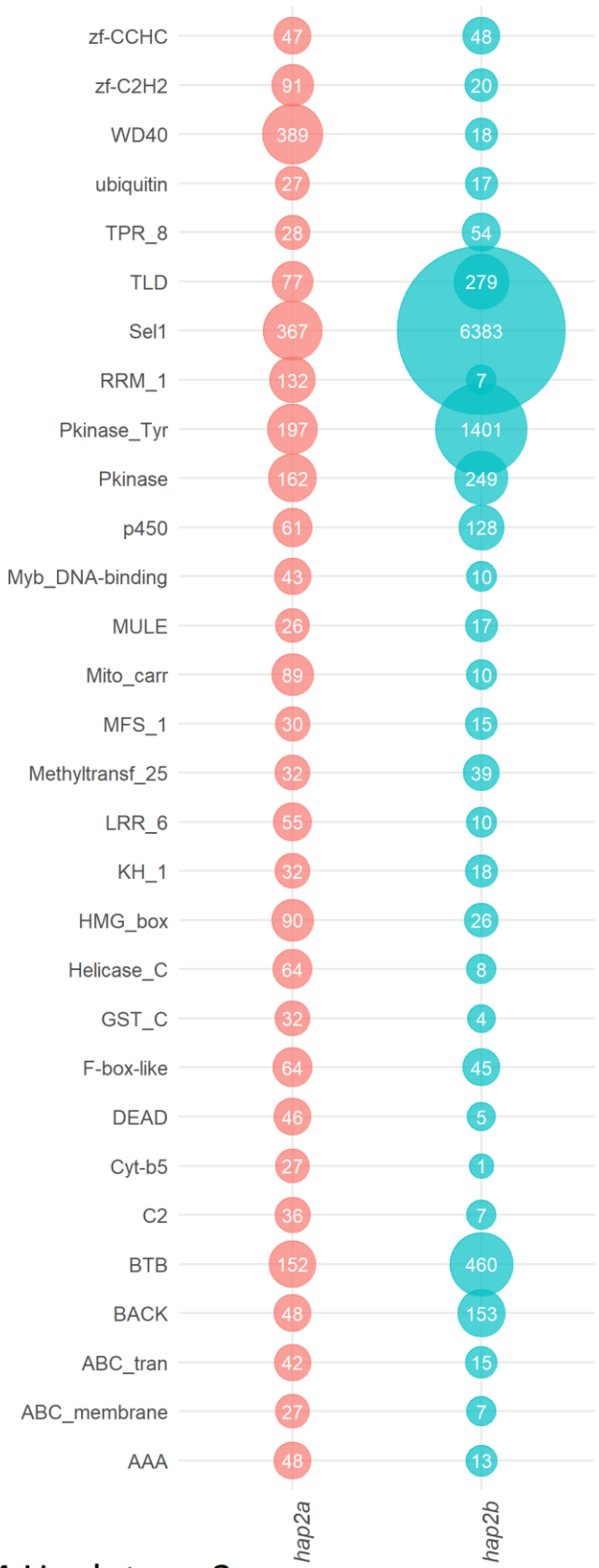

a5.2.chr1
a5.2.chr2
a5.2.chr3
a5.2.chr4
a5.2.chr5
a5.2.chr6
a5.2.chr7
a5.2.chr8
a5.2.chr9
a5.2.chr10
a5.2.chr11
a5.2.chr12
a5.2.chr13
a5.2.chr14
a5.2.chr15
a5.2.chr16
a5.2.chr17
a5.2.chr18
a5.2.chr19
a5.2.chr20
a5.2.chr21
a5.2.chr22
a5.2.chr23
a5.2.chr24
a5.2.chr25
a5.2.chr26
a5.2.chr27
a5.2.chr28
a5.2.chr29
a5.2.chr30
a5.2.chr31
a5.2.chr32
0b
1000Kb
2Mb
3Mb
4Mb
5Mb
6Mb
7Mb

Figure S3.3: The haplotypes of AMF dikaryons show strong chromosome compartmentalization. Red portions of the Karyoplots represents compartment A, and blue color represents compartment B.

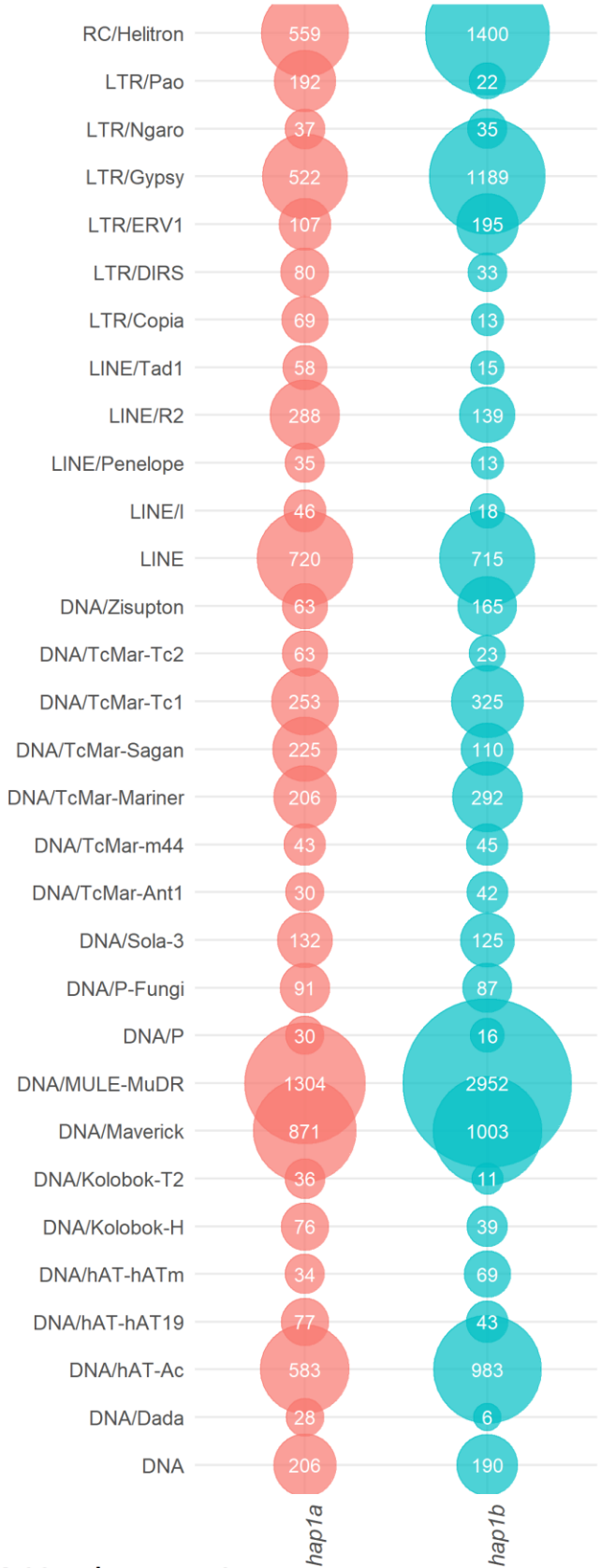


A4 Haplotype 1

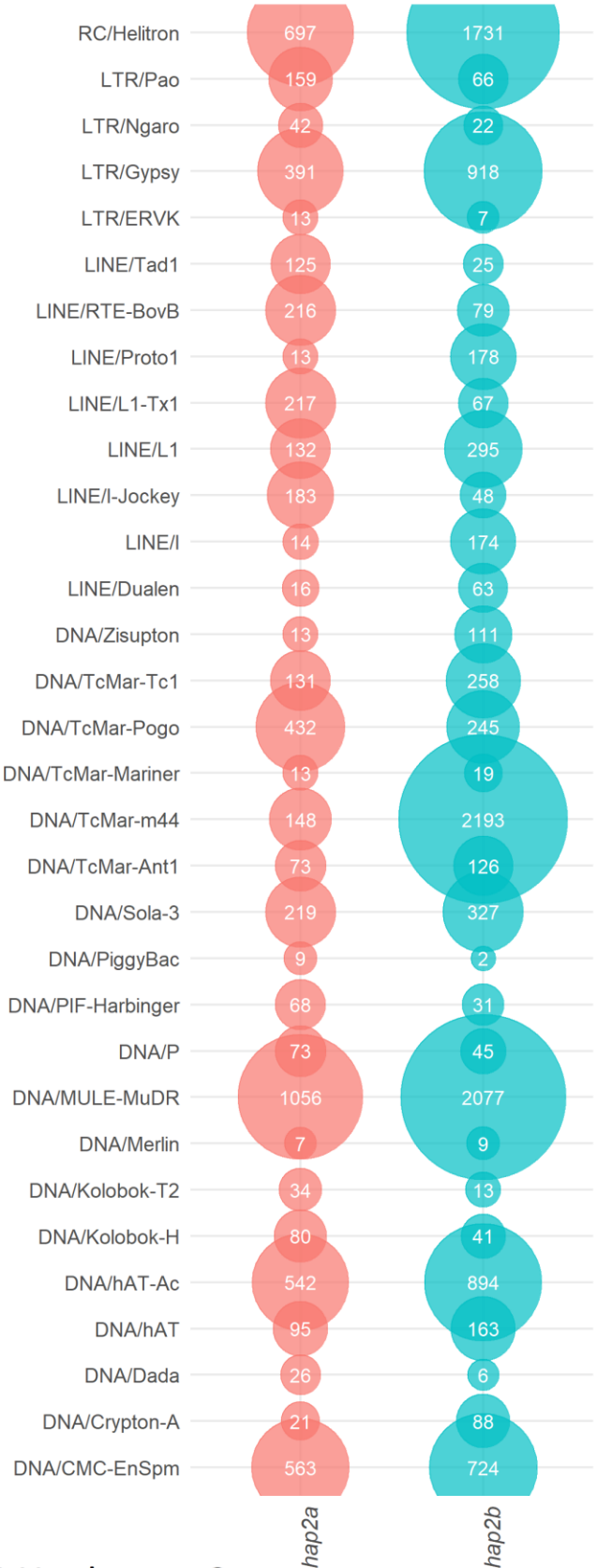


A4 Haplotype 2

Figure S3.4a: Bubbleplots representing the 30 most abundant Pfam domains of dikaryotic strain A4 for both haplotypes. The left column represents compartment A, and right column represents compartment B.



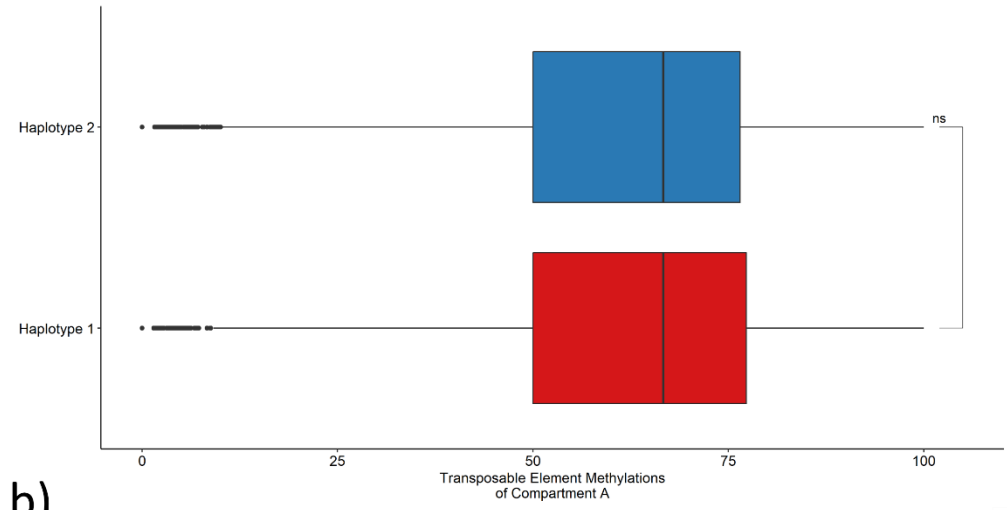
A4 Haplotype 1



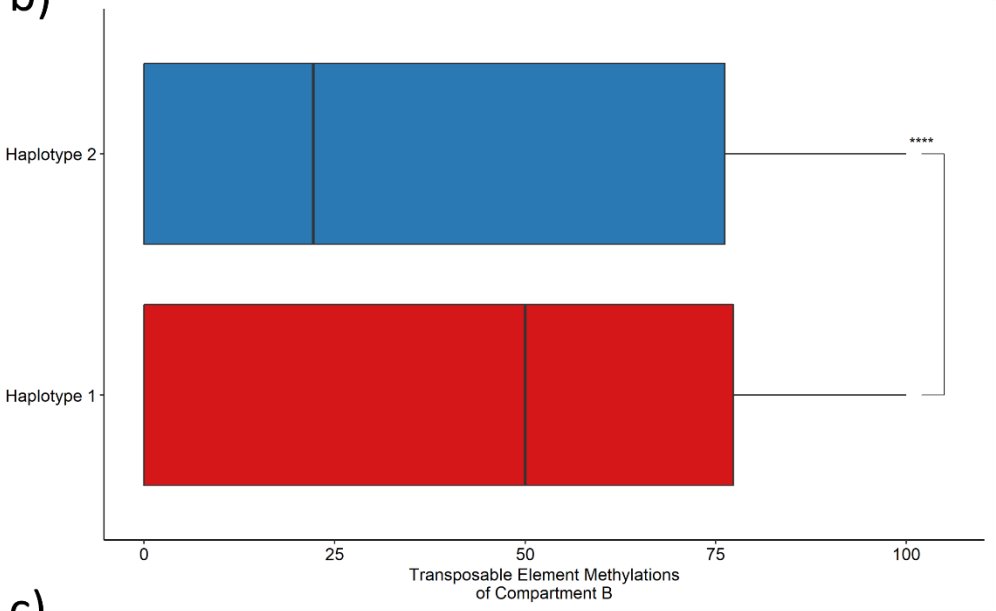
A4 Haplotype 2

Figure S3.4b: Bubbleplots representing the 30 most abundant repetitive elements of dikaryotic strains A4 in each haplotype. The left column represents compartment A, and right column represents compartment B.

a)


Haplotype 2
Haplotype 1
ns
0
25
50
75
100
Transposable Element Methylations
of Compartment A

b)


Haplotype 2
Haplotype 1
****
0
25
50
75
100
Transposable Element Methylations
of Compartment B

c)

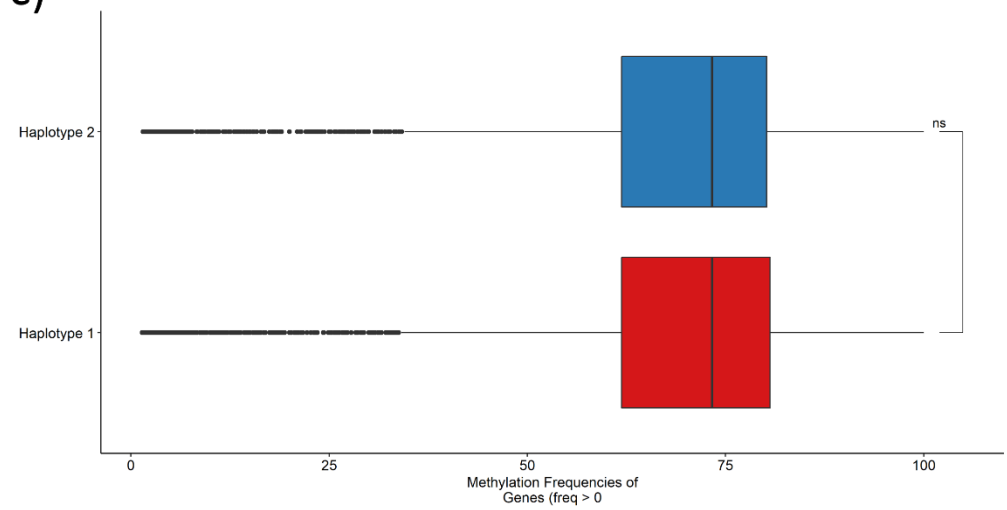
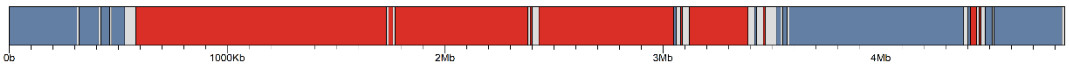

Haplotype 2
Haplotype 1
ns
0
25
50
75
100
Methylation Frequencies of
Genes (freq > 0

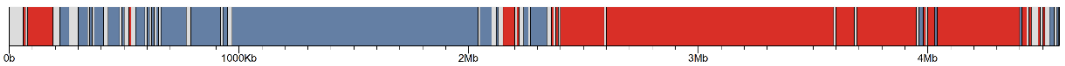
Figure S3.4c: Boxplots showing differences in methylation frequencies for each haplotype in dikaryotic strain A4. x-axis represents the methylation frequencies and y-axis represents the haplotypes. A) The difference in transposable element methylations in compartment A is not significant. B) Transposable elements found in B compartment of Haplotype 1 are methylated significantly heavier than transposable elements found in B compartment of Haplotype 2. (t-test, $p<0.0001$) C) After unmethylated bases are filtered out, methylation frequencies of genes located on both haplotypes do not differ significantly.

A4 Haplotype 1 chr19

A4 Haplotype 2 chr 19



A5 Haplotype 1 chr18



A5 Haplotype 2 chr18

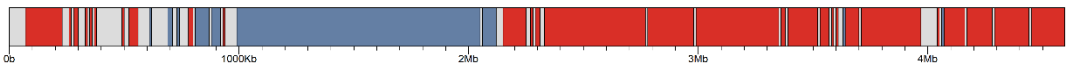
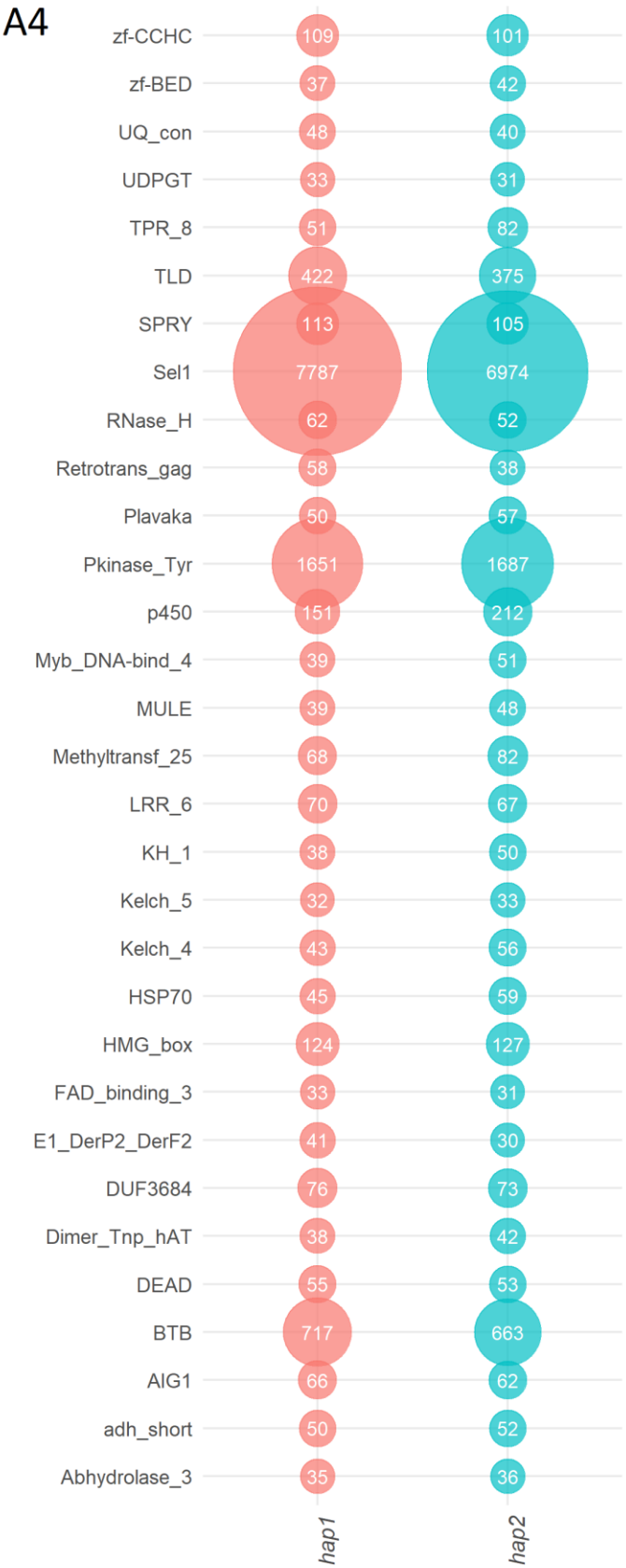
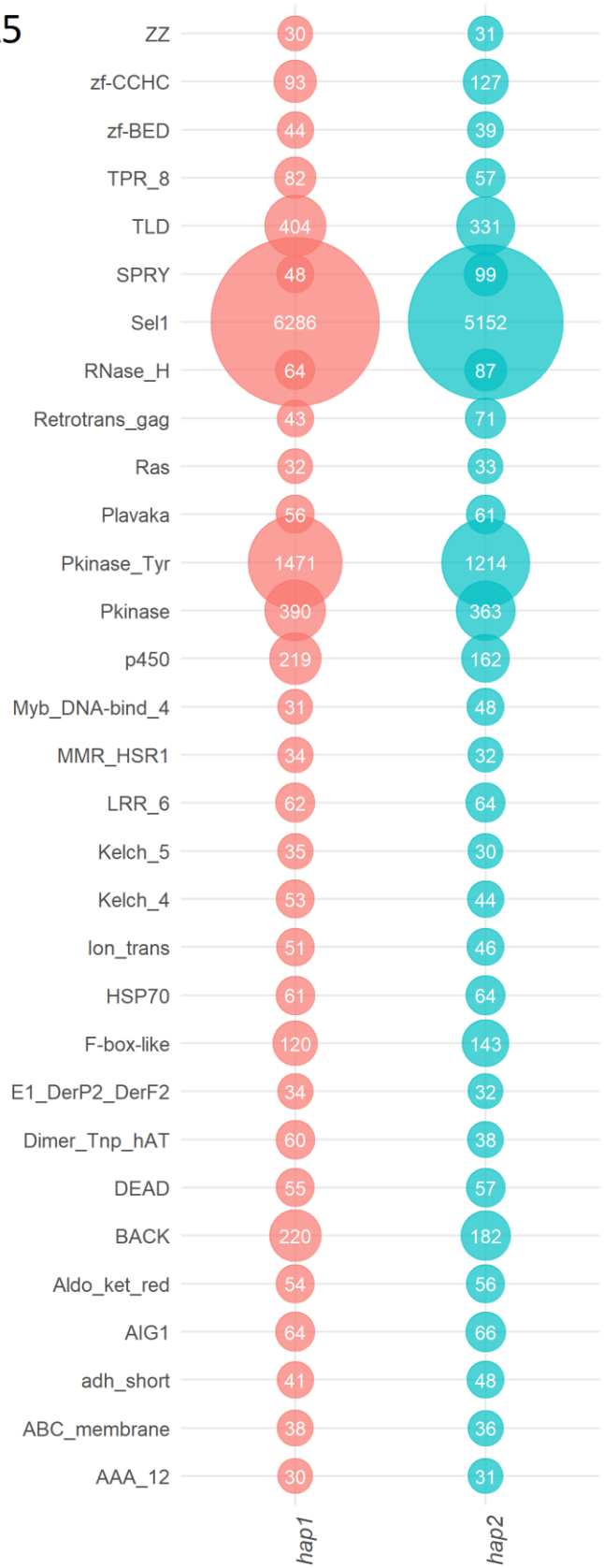


Figure S3.5: Compartmentalization of homologous chromosomes belonging to different parental haplotypes differs in certain chromosomes. Karyoplot colors represent compartments. Compartment A is shown in red color, and compartment B is shown in blue color.

A4


zf-CCHC
zf-BED
UQ_con
UDPGT
TPR_8
TLD
SPRY
Sel1
RNase_H
Retrotrans_gag
Plavaka
Pkinase_Tyr
p450
Myb_DNA-bind_4
MULE
Methyltransf_25
LRR_6
KH_1
Kelch_5
Kelch_4
HSP70
HMG_box
FAD_binding_3
E1_DerP2_DerF2
DUF3684
Dimer_Tnp_hAT
DEAD
BTB
AIG1
adh_short
Abhydrolase_3
109
37
48
33
51
422
113
7787
62
58
50
1651
151
39
39
68
70
38
32
43
45
124
33
41
76
38
55
717
66
50
35
101
42
40
31
82
375
105
6974
52
38
57
1687
212
51
48
82
67
50
33
56
59
127
31
30
73
42
53
663
62
52
36
hap1
hap2

A5


ZZ 30 31
zf-CCHC 93 127
zf-BED 44 39
TPR_8 82 57
TLD 404 331
SPRY 48 99
Sel1 6286 5152
RNase_H 64 87
Retrotrans_gag 43 71
Ras 32 33
Plavaka 56 61
Pkinase_Tyr 1471 1214
Pkinase 390 363
p450 219 162
Myb_DNA-bind_4 31 48
MMR_HSR1 34 32
LRR_6 62 64
Kelch_5 35 30
Kelch_4 53 44
Ion_trans 51 46
HSP70 61 64
F-box-like 120 143
E1_DerP2_DerF2 34 32
Dimer_Tnp_hAT 60 38
DEAD 55 57
BACK 220 182
Aldo_ket_red 54 56
AIG1 64 66
adh_short 41 48
ABC_membrane 38 36
AAA_12 30 31
hap1
hap2

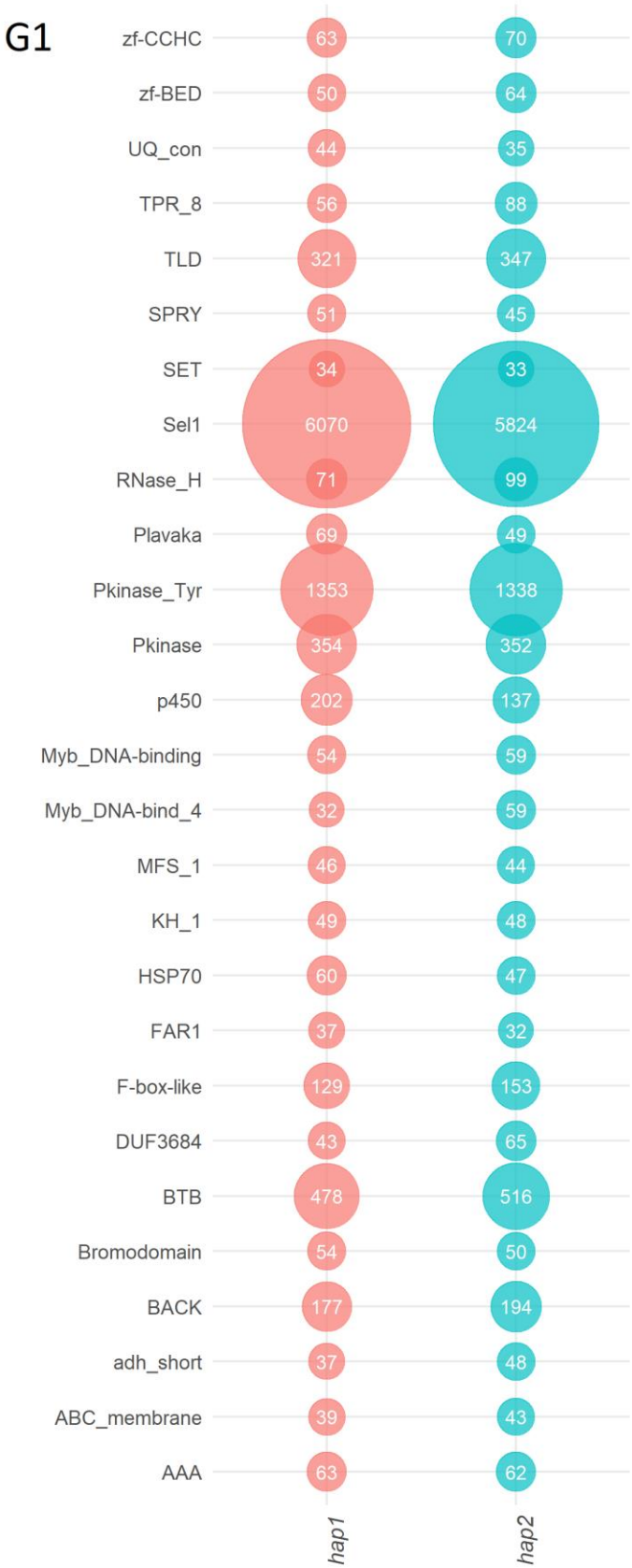
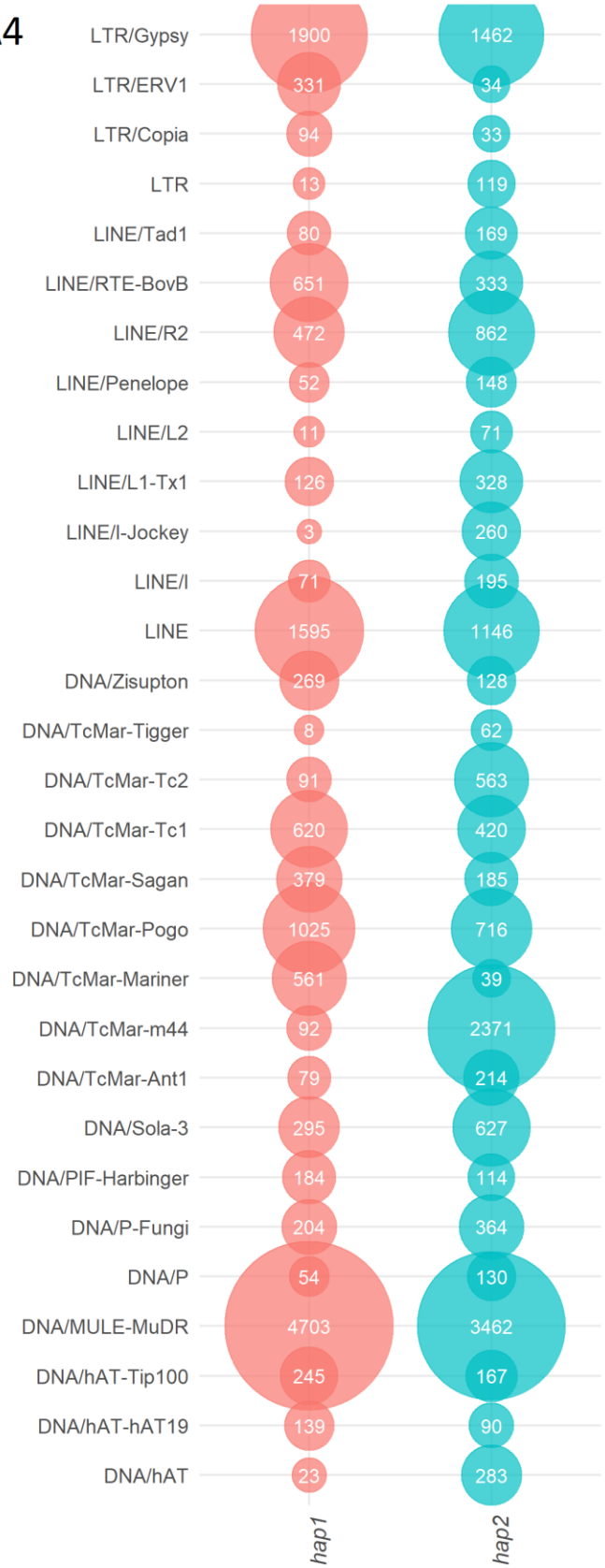
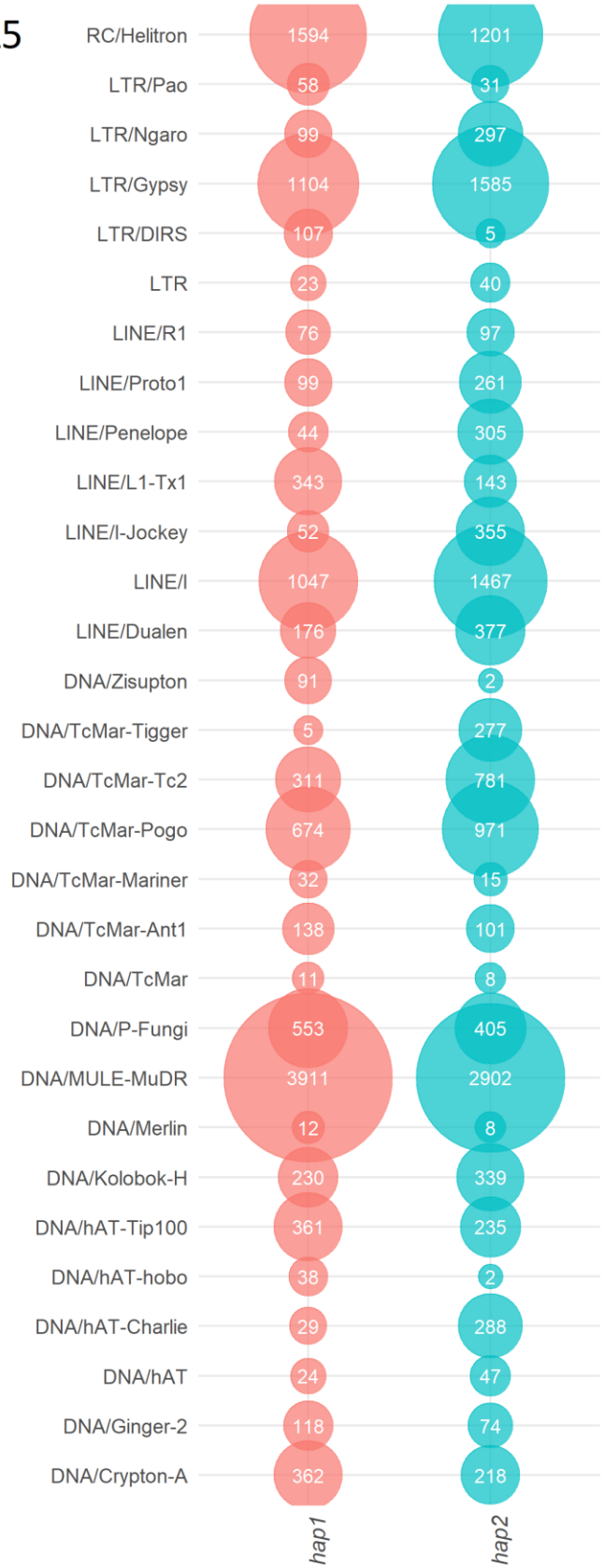

G1
zf-CCHC
zf-BED
UQ_con
TPR_8
TLD
SPRY
SET
Sel1
RNase_H
Plavaka
Pkinase_Tyr
Pkinase
p450
Myb_DNA-binding
Myb_DNA-bind_4
MFS_1
KH_1
HSP70
FAR1
F-box-like
DUF3684
BTB
Bromodomain
BACK
adh_short
ABC_membrane
AAA
63
50
44
56
321
51
34
6070
71
69
1353
354
202
54
32
46
49
60
37
129
43
478
54
177
37
39
63
70
64
35
88
347
45
33
5824
99
49
1338
352
137
59
59
44
48
47
32
153
65
516
50
194
48
43
62
hap1
hap2

Figure S3.6a: Bubbleplots representing the 30 most abundant Pfam domains of dikaryotic strain A4, A5 and G1 for each haplotype. The left column represents the Haplotype 1, and right column represents Haplotype 2.

A4


LTR/Gypsy
LTR/ERV1
LTR/Copia
LTR
LINE/Tad1
LINE/RTE-BovB
LINE/R2
LINE/Penelope
LINE/L2
LINE/L1-Tx1
LINE/I-Jockey
LINE/I
LINE
DNA/Zisupton
DNA/TcMar-Tigger
DNA/TcMar-Tc2
DNA/TcMar-Tc1
DNA/TcMar-Sagan
DNA/TcMar-Pogo
DNA/TcMar-Mariner
DNA/TcMar-m44
DNA/TcMar-Ant1
DNA/Sola-3
DNA/PIF-Harbinger
DNA/P-Fungi
DNA/P
DNA/MULE-MuDR
DNA/hAT-Tip100
DNA/hAT-hAT19
DNA/hAT
1900
331
94
13
80
651
472
52
11
126
3
71
1595
269
8
91
620
379
1025
561
92
79
295
184
204
54
4703
245
139
23
1462
34
33
119
169
333
862
148
71
328
260
195
1146
128
62
563
420
185
716
39
2371
214
627
114
364
130
3462
167
90
283
hap1
hap2

A5


RC/Helitron
LTR/Pao
LTR/Ngaro
LTR/Gypsy
LTR/DIRS
LTR
LINE/R1
LINE/Proto1
LINE/Penelope
LINE/L1-Tx1
LINE/I-Jockey
LINE/I
LINE/Dualen
DNA/Zisupton
DNA/TcMar-Tigger
DNA/TcMar-Tc2
DNA/TcMar-Pogo
DNA/TcMar-Mariner
DNA/TcMar-Ant1
DNA/TcMar
DNA/P-Fungi
DNA/MULE-MuDR
DNA/Merlin
DNA/Kolobok-H
DNA/hAT-Tip100
DNA/hAT-hobo
DNA/hAT-Charlie
DNA/hAT
DNA/Ginger-2
DNA/Crypton-A
1594
58
99
1104
107
23
76
99
44
343
52
1047
176
91
5
311
674
32
138
11
553
3911
12
230
361
38
29
24
118
362
1201
31
297
1585
5
40
97
261
305
143
355
1467
377
2
277
781
971
15
101
8
405
2902
8
339
235
2
288
47
74
218
hap1
hap2

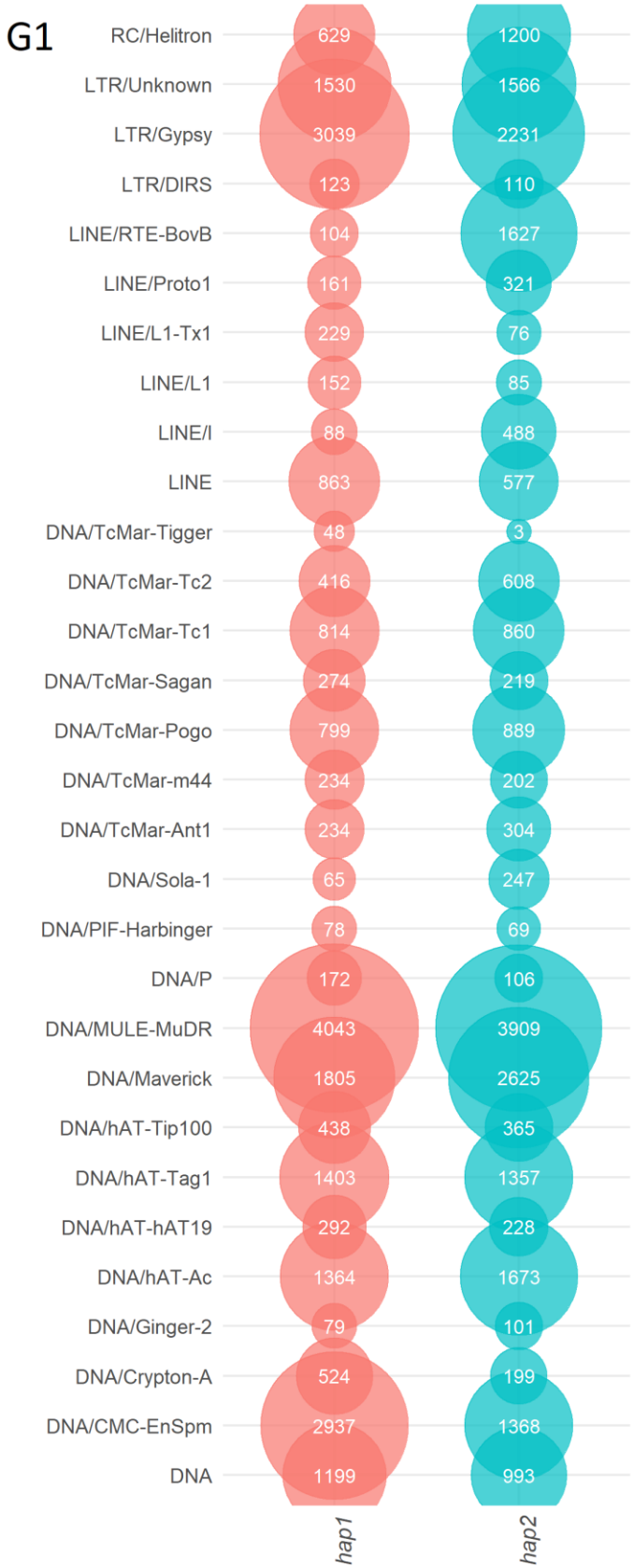
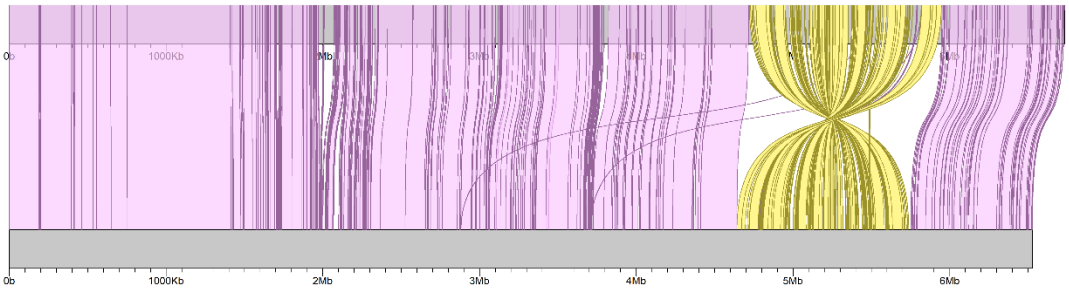

G1
RC/Helitron
LTR/Unknown
LTR/Gypsy
LTR/DIRS
LINE/RTE-BovB
LINE/Proto1
LINE/L1-Tx1
LINE/L1
LINE/I
LINE
DNA/TcMar-Tigger
DNA/TcMar-Tc2
DNA/TcMar-Tc1
DNA/TcMar-Sagan
DNA/TcMar-Pogo
DNA/TcMar-m44
DNA/TcMar-Ant1
DNA/Sola-1
DNA/PIF-Harbinger
DNA/P
DNA/MULE-MuDR
DNA/Maverick
DNA/hAT-Tip100
DNA/hAT-Tag1
DNA/hAT-hAT19
DNA/hAT-Ac
DNA/Ginger-2
DNA/Crypton-A
DNA/CMC-EnSpm
DNA
629
1530
3039
123
104
161
229
152
88
863
48
416
814
274
799
234
234
65
78
172
4043
1805
438
1403
292
1364
79
524
2937
1199
1200
1566
2231
110
1627
321
76
85
488
577
3
608
860
219
889
202
304
247
69
106
3909
2625
365
1357
228
1673
101
199
1368
993
hap1
hap2

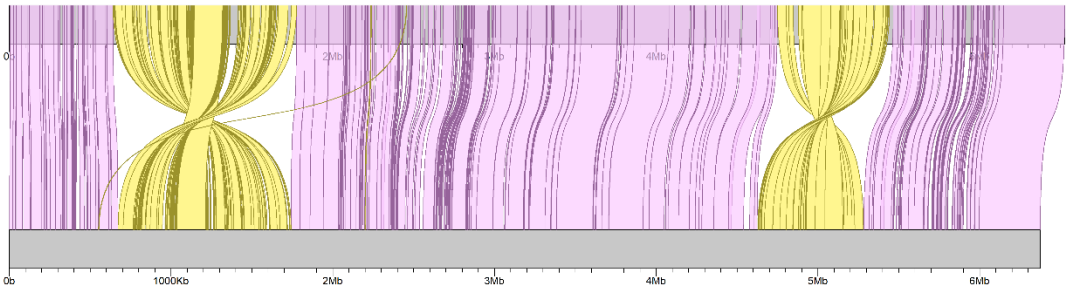
Figure S3.6b: Bubbleplots representing the 30 most abundant repetitive elements of dikaryotic strains A4, A5 and G1 for each haplotype. The left column represents the Haplotype 1, and right column represents Haplotype 2.

A4 Haplotype 1 chr4



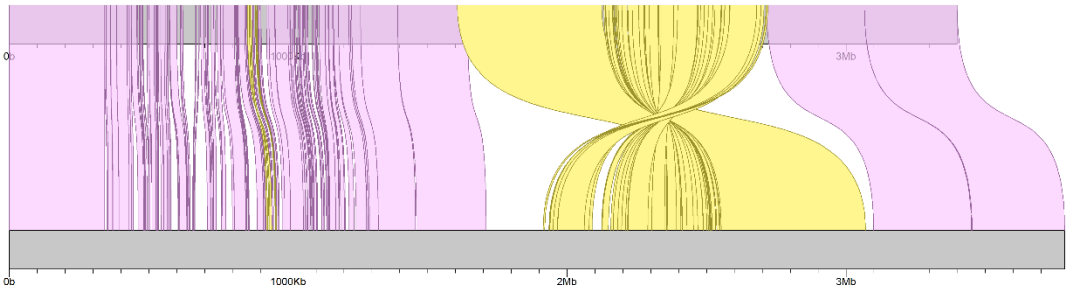
A4 Haplotype 2 chr4

A5 Haplotype 1 chr3



A5 Haplotype 2 chr3

G1 Haplotype 1 chr23



G1 Haplotype 2 chr23

Figure S3.6c: Rearrangement events between homologous chromosomes of parental haplotypes. Synteny blocks are shown in pink, and inversion events are shown in yellow.

Table S3.1: Operon locations for all parental haplotypes

Assigned operon ID	Location	AMF Strain
operon 1	chr23_A4_Hap1:3948972-3954773	A4 hap1
operon 2	chr23_A4_Hap1:4033538-4027737	A4 hap1
operon 3	chr17_A4_Hap1:1468375-1462560	A4 hap1
operon 4	chr18_A4_Hap1:1646399-1640580	A4 hap1
operon 5	chr18_A4_Hap1:2012308-2006481	A4 hap1
operon 6	chr18_A4_Hap1:1957160-1962989	A4 hap1
operon 7	chr18_A4_Hap1:1851467-1857283	A4 hap1
operon8	chr2_A4_Hap1:6786257-6792080	A4 hap1
operon9	chr9_A4_Hap1:1106656-1100852	A4 hap1
operon1	chr23_A4_Hap2:3825018- 3830819	A4 hap2
operon2	chr23_A4_Hap2:3909584-3903783	A4 hap2
operon3	chr17_A4_Hap2:1439838-1434023	A4 hap2
operon4	chr18_A4_Hap2:1660009 -1654191	A4 hap2
operon5	chr18_A4_Hap2:1818184-1823998	A4 hap2
operon6	chr18_A4_Hap2:1951508-1957335	A4 hap2
operon7	chr18_A4_Hap2:2017679-2011852	A4 hap2
operon8	chr2_A4_Hap2:6696150-6701973	A4 hap2
operon9	chr9_A4_Hap2:1179831-1174028	A4 hap2
operon1	chr23_A5_Hap1:3040099-3045898	A5 hap1
operon2	chr23_A5_Hap1:3217915-3212116	A5 hap1

operon3	chr18_A5_Hap1:1764572-1770392	A5 hap1
operon4	chr18_A5_Hap1:1868777-1862957	A5 hap1
operon5	chr18_A5_Hap1:1462146-1467979	A5 hap1
operon6	chr18_A5_Hap1:1664019-1658186	A5 hap1
operon7	chr18_A5_Hap1:1271000-1265167	A5 hap1
operon8	chr17_A5_Hap1:1415295-1409480	A5 hap1
operon9	chr2_A5_Hap1:7017538-7011709	A5 hap1
operon10	chr9_A5_Hap1:928574-922772	A5 hap1
operon1	chr23_A5_Hap2:3117778-3123577	A5 hap2
operon2	chr23_A5_Hap2:3281085-3275286	A5 hap2
operon3	chr18_A5_Hap2:1773370-1779190	A5 hap2
operon4	chr18_A5_Hap2:1877575-1871755	A5 hap2
operon5	chr18_A5_Hap2:1470943-1476776	A5 hap2
operon6	chr18_A5_Hap2:1672817-1666984	A5 hap2
operon7	chr18_A5_Hap2:1277461-1271628	A5 hap2
operon8	chr17_A5_Hap2:1326027-1320212	A5 hap2
operon0	chr2_A5_Hap2:6882018-6876189	A5 hap2
operon10	chr9_A5_Hap2:840101-834299	A5 hap2
operon1	chr23_G1_Hap1:2965539-2971337	G1 hap1
operon2	chr23_G1_Hap1:3066486-3060688	G1 hap1
operon3	chr18_G1_Hap1:1244037-1238217	G1 hap1
operon4	chr18_G1_Hap1:1445816-1451637	G1 hap1
operon5	chr18_G1_Hap1:1557047-1551226	G1 hap1

operon6	chr18_G1_Hap1:1631553-1637381	G1 hap1
operon7	chr18_G1_Hap1:1877466-1871638	G1 hap1
operon8	chr17_G1_Hap1:1315313-1309495	G1 hap1
operon9	chr17_G1_Hap1:1343127-1337093	G1 hap1
operon10	chr2_G1_Hap1:5632794-5638615	G1 hap1
operon11	chr9_G1_Hap1:1019906-1014104	G1 hap1
operon1	chr23_G1_Hap2:3348848-3354646	G1 hap2
operon2	chr23_G1_Hap2:3452200-3443996	G1 hap2
operon3	chr17_G1_Hap2:1316163-1310345	G1 hap2
operon4	chr18_G1_Hap2:1199255-1193435	G1 hap2
operon5	chr18_G1_Hap2:1401034-1406855	G1 hap2
operon6	chr18_G1_Hap2:1512240-1506419	G1 hap2
operon7	chr18_G1_Hap2:1586746-1592574	G1 hap2
operon8	chr18_G1_Hap2:1832659-1826831	G1 hap2
operon9	chr2_G1_Hap2:5622290-5628111	G1 hap2
operon10	chr9_G1_Hap2:1116432-1110629	G1 hap2

Table S3.2: Functional classification of eukaryotic (KOG) clusters of haplotype-specific genes.

	Abundance					
	A4		A5		G1	
KOG description	**Hap 1**	**Hap 2**	**Hap 1**	**Hap 2**	**Hap 1**	**Hap 2**
Translation, ribosomal structure and biogenesis	0.019572	0.018221	0.015085	0.048313	0.03953	0.040584
RNA processing and modification	0.016407	0.008349	0.013308	0.024441	0.025093	0.027257
Transcription	0.026611	0.032842	0.042612	0.044871	0.03656	0.048502
Replication, recombination and repair	0.04437	0.037659	0.020398	0.030287	0.033565	0.034879
Chromatin structure and dynamics	0.013825	0.010977	0.004062	0.001815	0.018661	0.017532
Cell cycle control, cell division, chromosome partitioning	0.013947	0.018059	0.026464	0.029378	0.027995	0.027563
Nuclear structure	0.001221	0.004414	0.002252	0.00207	0.003129	0.002369
Defense mechanisms	0.008142	0.005049	0.010489	0.002156	0.008739	0.015165
Signal transduction mechanisms	0.597382	0.620048	0.556926	0.473862	0.371647	0.349558
Cell wall/membrane/envelope biogenesis	0.224589	0.146484	0.115229	0.093562	0.120967	0.094142
Cell motility	0	0	0	0	0.000794	0
Cytoskeleton	0.013684	0.014574	0.009302	0.021625	0.023616	0.017402
Extracellular structures	0.000167	0.000145	0.004955	0	0.001597	0.001575
Intracellular trafficking, secretion, and vesicular transport	0.016434	0.011975	0.017616	0.026162	0.041106	0.033514
Posttranslational modification, protein turnover, chaperones	0.28837	0.187028	0.156862	0.19381	0.186826	0.172424
Energy production and conversion	0.009839	0.024222	0.037518	0.025869	0.038354	0.034462
Carbohydrate transport and metabolism	0.004821	0.001414	0.028611	0.003686	0.027135	0.02547
Amino acid transport and metabolism	0.015329	0.012727	0.015738	0.021608	0.028812	0.019722
Nucleotide transport and metabolism	0.002911	0.00243	0.012664	0	0.007535	0.007837
Coenzyme transport and metabolism	0.00269	0.006789	0.012496	0	0.007897	0.008729
Lipid transport and metabolism	0.0272	0.05383	0.069097	0.039954	0.057126	0.054919
Inorganic ion transport and metabolism	0.011711	0.005882	0.011987	0.007303	0.016979	0.017383
Secondary metabolites biosynthesis, transport and catabolism	0.048304	0.063568	0.081071	0.029485	0.041617	0.031883
General function prediction only	0.072794	0.069343	0.07152	0.145871	0.127377	0.138469
Function unknown	0.036277	0.027816	0.025914	0.022935	0.041605	0.05847

Conclusion

The fossil evidence shows that arbuscular mycorrhizal fungi have emerged hundreds of millions of years ago (Remy *et al.*, 1994; Redecker *et al.*, 2000) and it is known that these organisms can be found in all continents, in symbiosis with their plant hosts (Barbosa *et al.*, 2017). Despite their evolutionary and ecological success, AMF were classified as "evolutionary scandals", or "ancient asexuals" for decades, because no sexual reproductive structures were formally observed in these fungi. With a strictly clonal reproduction system, how AMF can achieve this major ecological and evolutionary success has long puzzled scientists and this notion has been a focus in fungal research for many years.

One of the earlier hypotheses that aim for explaining this success was the "heterokaryosis" hypothesis. This hypothesis proposed that AMF contain thousands of highly diverging and recombining nuclei in their cytoplasm (Ehinger *et al.*, 2009; Boon *et al.*, 2015), creating an intrinsic and perpetual genetic diversity that allows for adaptability in the absence of sex. Although this theory was predominant for over a decade, it was severely challenged by other studies showing evidence of genetic homogeneity among nuclei (Pawlowska & Taylor, 2005), and by whole genome sequencing studies conducted on *Rhizophagus irregularis*, a model AMF species, which revealed no evidence of unusual nuclear variability within the mycelium (Tisserant *et al.*, 2013; Lin *et al.*, 2014).

More recently, the discovery of putative MAT loci and meiosis genes in the genomes of *R. irregularis* lead to the emerging view that these organisms may have a cryptic sexual (or parasexual) life cycle. This view is strongly supported by evidence that model AMF strains are either homokaryotic or dikaryotic (Ropars *et al.*, 2016), mirroring life-stages seen in other fungi,

particularly Basidiomycetes. Even though this view has resulted in paradigm shifts in our understanding of AMF genetics, it was also built on highly fragmented datasets. As such, to fully understand the genetics and genome biology of AMF, high quality reference genome assemblies from representatives of AMF homokaryons and dikaryons were still necessary.

The multinucleated nature of AMF combined with the high repeat content of their genomes prevented us obtain a detailed view of their genome content and structure. Short sequencing reads produced by the Illumina platform has resulted in highly fragmented genome assemblies. These genome assemblies can contain many artefacts such as collapsed and duplicated regions and chimeric assemblies (Denton *et al.*, 2014; Montoliu-Nerin *et al.*, 2020). The high fragmentation rate also prevented in-depth and reliable analyses of the AMF genome biology and epigenetics.

In this PhD thesis I tackled this by producing and analyzing chromosomal level genome assemblies, resulting in significant insights on the genome content, epigenetics and mode of evolution of five strains of the AMF model species *Rhizophagus irregularis.* These genome assemblies and the analyses I performed with collaborators lay the foundation for genome research in AMF, and more generally for research that focuses on genome analyses of mutualistic fungal species and host-microbe interactions.

Summary of main findings

The first chapter of this PhD thesis reviews the possible causes of increased AMF nuclear diversity (Yildirir *et al.*, 2020). Even though evidence for nuclear recombination was found in AMF genomes through whole genome analysis studies (Vandenkoornhuyse *et al.*, 2001; Chen *et al.*, 2018a), sexual reproduction has never been observed. In this review, the parasexual mechanisms have been proposed as a possible driving force for the high genome diversity in AMF strains. As it is established that the transposable elements have the power to trigger the genetic transfer of protein-coding genes across species in pathogenic fungi (McDonald *et al.*, 2019), I also proposed a similar mechanism in AMF where the high repeat content of the AMF genomes might lead to somatic recombination events between nuclei, but empirical testing is required to further support this hypothesis.

My second and third chapters focused on constructing reference level genome assemblies for homokaryotic and dikaryotic AMF strains and analyzing their genome content and biology in detail for the first time. For both chapters, I used single molecule sequencing platforms and chromatin capture technologies to construct chromosomal genome assemblies and identify the chromosome confirmation of these strains. The new genome assemblies constructed in my thesis surpass the previous genome assembly studies of *R. irregularis* (Maeda *et al.*, 2018b; Chen *et al.*, 2018a) in every way. Most of the chromosomes contain two telomeres, the repeat regions are captured in full and gene annotations and BUSCO scores are greatly improved compared to the previous work on this species. In particular, the use of Hi-C reads helped correcting genome artefacts (chimeric/collapsed/incorrect join regions) that would normally go unnoticed in other studies were corrected, resulting in the best genome assemblies produced to date.

In my second chapter, I produced five homokaryotic *R. irregularis* genome assemblies, analyzed their genome content and was able to discover the two-compartmental nature of their chromosomal folding. This compartmentalization has a direct influence on AMF genome biology and evolution. Specifically, the compartment A contains most of the housekeeping genes, has a higher average transcription and overall higher transposable element methylation frequencies. On the other hand, compartment B contains most of the unique genes, has a lower transcription profile and the transposable elements located there do not have strong methylation. This compartment also rearranges at a higher rate.

In my third and final chapter, I focused on understanding the genome content and biology of dikaryotic *R. irregularis* strains. To this end, I successfully assembled their parental haplotypes at the chromosome-level. Analysis of parental haplotypes revealed their unique genetic content and evolution, and thus their putative respective role in the biology of these genetically peculiar strains.

Future Directions

Origin of recombination in AMF

AMF genome assemblies and single nucleus data carry many signatures of recombination events (Vandenkoornhuyse *et al.*, 2001; Chen *et al.*, 2018a), signaling enhanced genetic diversity of the strains induced by parasexual reproduction events. The new analyses performed on the parental assemblies of *R. irregularis* confirmed these results; AMF dikaryons carry evidence of somatic recombination within their mycelium as a result of genetic exchange between nuclei of different mating type.

Even though several proposed sexual and parasexual reproduction models might explain this enhanced genetic diversity, experimental testing of these models is required to conclusively confirm their presence and to determine the frequency of them. Now that complete genome assemblies are available, such analyses can be performed with great detail. One such analysis could be done through crossing experiments. For example, after successful crossing of two compatible homokaryotic strains, the genome content of new single spores isolated from those crossed strains could be investigated using molecular approaches with genetic marker specific to each parental and designed on their respective chromosome data. In this case, the co-existence of genotypes from crossed dikaryons would provide compelling evidence of recombination between those strains, presumably because of sexual reproduction.

In addition to that, to discover additional parasexual events, we can analyze genetically divergent nuclei found in dikaryotic strains using long-read data, as opposed to short paired-end reads. The combination of complete parental haplotypes from dikaryotic strains, and long-read information from individual nuclei should reveal conclusive and detailed evidence on the

existence and frequency of inter-nucleus recombination in these strains. It would provide a detailed view of recombinant breaks and the role of transposable regions in creating this genetic exchange.

AMF Genome Biology, Epigenetics and non-model AMF Species

Overall, my thesis contributes significantly to the understanding of AMF genomics, but also generates many new questions. One of these questions arises from the second chapter of my thesis when we discuss the compartmentalization of the genomes. In this chapter, we were able to determine that compartments differ in overall gene expression; compartment A is transcriptionally active whereas compartment B shows the characteristics of a heterochromatin region. However, this difference in transcription levels is reversed when the *in planta* secreted protein transcription is analyzed.

In planta, compartment B's transcription levels are upregulated for some secreted genes/effectors, but this transcription increase is not observed in compartment A. We hypothesised that this change in the transcription levels results from a change in genome organization affecting the compartment B. Specifically, we proposed that molecular dialogues with the plant host induces a relaxation of the B compartment, leading to upregulation of specific genes involved in the dialogue between partner, such as effectors. It is possible that other external factors such as different plant hosts, various environmental stress conditions, different life stages may lead to similar changes in genome organization. This hypothesis suggests that the chromosome organization of AMF is not static, but it changes depending on the surrounding environment and possibly the host identity. Such changes are seen in yeast genomes (Kim *et al.*,

2017) and along life-stages in model eukaryotes (Falk *et al.*, 2019; Matthey-Doret *et al.*, 2022), but have not been tested empirically in AMF. To test this hypothesis, future work should obtain Hi-C data from in-planta AMF material, as opposed to extra-radical mycelium.

In this case, a change in the A and B compartments would support the notion that these compartments are malleable, and this data should be supported by RNA-seq data to determine how this epigenetic variability affects the transcriptome. Furthermore, in- planta Hi-C data from multiple hosts could also reveal if the AMF genome biology, including chromosome conformation and gene expression, change depending on the host identify. Similar analyses could also focus on cultures produced under different conditions, including stress (pH, temperature etc.). This work would ultimately reveal a new layer of complexity in the interactions between partners of the mycorrhizal symbiosis, and across conditions.

Moreover, even though DNA methylations and chromosomal compartments were analyzed and a direct link to gene transcriptions is found in this thesis, to fully reveal the epigenetic landscape of AMF and its link to the transcription levels, we must also consider the histone modifications. By combining DNA methylations, histone modifications and chromosomal compartment organization of the genomes, we can identify mechanisms leading to the increased genetic and epigenetic variability we observed between strains. This would help us further investigate the biological changes AMF genomes go through to support plant hosts in various environmental conditions.

In this thesis, I focused mainly on one species, *Rhizophagus irregularis*, to study AMF genome content and biology. This is because this species grows well in root organ cultures, meaning sufficient DNA and RNA can be obtained for high quality genome work, and the species represents the model used by most AMF researchers. Ultimately, to acquire an

understanding of AMF biology beyond model species, the analyses we produced should be produced on all members of this subphylum and, ideally, to multiple strains of the same species across AMF communities and ecosystems. This will be essential to better our understanding of this prominent terrestrial symbiosis.

Lastly, one question that my thesis could not answer is regarding the centromeric regions of AMF genomes. Other fungal genomes either contain "regional centromeres", long stretches of repetitive DNA in their centromeric region (Sanyal *et al.*, 2004; Cam *et al.*, 2005; Baum *et al.*, 2006), or "point centromeres", extremely short centromeres with specific DNA sequences (Meraldi *et al.*, 2006). AMF genomes do not contain obvious regional centromeres and as transcription is present throughout the chromosomes and repetitive regions are dispersed without an order, AMF centromeres most likely show the properties of point centromeres. To pinpoint the centromere locations, we must now first identify and study the AMF kinetochores – i.e. the location where chromosomes bind. This work will likely require the use of cutting-edge microscopy techniques and will ultimately be necessary to determine with exactitude how many chromosomes exist in *Rhizophagus irregularis*.

Function Of Parental Haplotypes in the Symbiosis and Origin of AMF Dikaryons

Another big question that comes from chapter three is how parental haplotype interact with the host and environmental conditions in dikaryotic strains. Specifically, my thesis showed that parental haplotypes differ significantly in gene content and function, but how these changes affect their respective contribution to the transcriptome of these strains is unknown. To address this, future studies should start analyzing haplotype transcriptome to reveal which genes each are

used depending on the host under different environmental and stress conditions. Analyzing the transcriptome data is an essential step in understanding AMF biology and it will help us identify the complex balance between the two haplotypes.

The constructed phylogenetic tree in chapter three revealed that parental haplotypes are much more similar to each other than other homokaryotic strains. This indicates that the AMF dikaryons we studied did not emerge through the hybridization of two genetically distinct homokaryotic strains, as seen for example in fungal pathogens (Li *et al.*, 2019), but they more likely emerged through a nuclear exchange event between two very similar homokaryotic strains. With this information in hand, the community has now the theoretical knowledge to find and cross genetically similar and sexually compatible homokaryotic strains. Specifically, our findings indicate that to successfully produce a dikaryotic progeny (or a recombined progeny), one should cross AMF homokaryons that are phylogenetically very similar but still carry different MAT-loci. To be able to do this, it is essential to increase the genomic sampling efforts and include more members of *Rhizophagus irregularis* and other AMF species to the phylogenetic tree to obtain a better understanding of genetic in this model species as means to find the most suitable candidates for crossing.

References

Anderson JB, Kohn LM. **2014**. Dikaryons, Diploids, and Evolution. *Sex in Fungi*: 333–348.

Anderson MZ, Thomson GJ, Hirakawa MP, Bennett RJ. **2019**. A 'parameiosis' drives depolyploidization and homologous recombination in Candida albicans. *Nature Communications* **10**: 4388.

Arima Genomics. **2019**. 01_mapping_arima.sh.

Arkhipova I, Meselson M. **2005**. Deleterious transposable elements and the extinction of asexuals. *BioEssays* **27**: 76–85.

Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. **2020**. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. *BMC Biology* **18**.

den Bakker HC, Vankuren NW, Morton JB, Pawlowska TE. **2010**. Clonality and Recombination in the Life History of an Asexual Arbuscular Mycorrhizal Fungus. *Molecular Biology and Evolution* **27**: 2474–2486.

Barbosa M v., Pereira EA, Cury JC, Carneiro MAC. **2017**. Occurrence of arbuscular mycorrhizal fungi on King George Island, South Shetland Islands, Antarctica. *Anais da Academia Brasileira de Ciências* **89**: 1737–1743.

Baum M, Sanyal K, Mishra PK, Thaler N, Carbon J. **2006**. Formation of functional centromeric chromatin is specified epigenetically in Candida albicans. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 14877–14882.

Bécard G, Pfeffer PE. **1993**. Status of nuclear division in arbuscular mycorrhizal fungi during in vitro development. *Protoplasma 1993 174:1* **174**: 62–68.

Bertazzoni S, Williams AH, Jones DA, Syme RA, Tan K-C, Hane JK. **2018**. Accessories Make the Outfit: Accessory Chromosomes and Other Dispensable DNA Regions in Plant-Pathogenic Fungi. *Molecular Plant-Microbe Interactions®* **31**: 779–788.

Bonfante P, Anca IA. **2009**. Plants, mycorrhizal fungi, and bacteria: A network of interactions. *Annual Review of Microbiology* **63**: 363–383.

Bonfante P, Genre A. **2010**. Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature Communications 2010 1:1* **1**: 1–11.

Boon E, Halary S, Bapteste E, Hijri M. **2015**. Studying Genome Heterogeneity within the Arbuscular Mycorrhizal Fungal Cytoplasm. *Genome Biology and Evolution* **7**: 505–521.

Bruns TD, Corradi N, Redecker D, Taylor JW, Öpik M. **2018**. Glomeromycotina: what is a species and why should we care? *New Phytologist* **220**: 963–967.

Buchfink B, Reuter K, Drost HG. **2021**. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods 2021 18:4* **18**: 366–368.

Buchfink B, Xie C, Huson DH. **2014**. Fast and sensitive protein alignment using DIAMOND. *Nature Methods 2014 12:1* **12**: 59–60.

Cabanettes F, Klopp C. **2018**. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**: e4958.

Cam HP, Sugiyama T, Chen ES, Chen X, FitzGerald PC, Grewal SIS. **2005**. Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature Genetics 2005 37:8* **37**: 809–819.

Carvalho J, Negrinho R, Azinheiro S, Garrido-Maestu A, Barros-Velázquez J, Prado M. **2018**. Novel approach for accurate minute DNA quantification on microvolumetric solutions. *Microchemical Journal* **138**: 540–549.

Casselton LA. **2008**. Fungal sex genes - Searching for the ancestors. *BioEssays* **30**: 711–714.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. **2013**. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–3140.

Chen ECH, Mathieu S, Hoffrichter A, Ropars J, Dreissig S, Fuchs J, Brachmann A, Corradi N. **2020**. More Filtering on SNP Calling Does Not Remove Evidence of Inter-Nucleus Recombination in Dikaryotic Arbuscular Mycorrhizal Fungi. *Frontiers in Plant Science* **11**: 912.

Chen ECH, Mathieu S, Hoffrichter A, Sedzielewska-Toro K, Peart M, Pelin A, Ndikumana S, Ropars J, Dreissig S, Fuchs J, *et al.* 2018a. Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. *eLife* **7**.

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

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. **2021**. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *NaTuRe MeTHods | **18***.

Clark TA, Anderson JB. **2004**. Dikaryons of the basidiomycete fungus Schizophyllum commune: evolution in long-term culture. *Genetics* **167**: 1663–1675.

Cornell C, Kokkoris V, Turcu B, Dettman J, Stefani F, Corradi N. **2022**. The arbuscular mycorrhizal fungus Rhizophagus irregularis harmonizes nuclear dynamics in the presence of distinct abiotic factors. *Fungal Genetics and Biology* **158**: 103639.

Corradi N, Bonfante P. **2012**. The Arbuscular Mycorrhizal Symbiosis: Origin and Evolution of a Beneficial Plant Infection. *PLoS Pathogens* **8**.

Corradi N, Brachmann A. **2017**. *Fungal Mating in the Most Widespread Plant Symbionts?* Elsevier.

Corradi N, Croll D, Colard A, Kuhn G, Ehinger M, Sanders IR. **2007**. Gene copy number polymorphisms in an arbuscular mycorrhizal fungal population. *Applied and Environmental Microbiology* **73**: 366–369.

Corradi N, Hijri M, Fumagalli L, Sanders IR. **2004a**. Arbuscular mycorrhizal fungi (Glomeromycota) harbour ancient fungal tubulin genes that resemble those of the chytrids (Chytridiomycota). *Fungal Genetics and Biology* **41**: 1037–1045.

Corradi N, Kuhn G, Sanders IR. **2004b**. Monophyly of β-tubulin and H+-ATPase gene variants in Glomus intraradices: consequences for molecular evolutionary studies of AM fungal genes. *Fungal Genetics and Biology* **41**: 262–273.

Croll D, McDonald BA. **2012**. The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. *PLOS Pathogens* **8**: e1002608.

Croll D, Sanders IR. **2009**. Recombination in Glomus intraradices, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC Evolutionary Biology* **9**: 1–11.

Croll D, Wille L, Gamper HA, Mathimaran N, Lammers PJ, Corradi N, Sanders IR. **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**: 672–687.

Dekker J, Heard E. **2015**. Structural and functional diversity of Topologically Associating Domains. *FEBS Letters* **589**: 2877–2884.

Delaux PM, Schornack S. **2021**. Plant evolution driven by interactions with symbiotic and pathogenic microbes. *Science* **371**.

Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, Hahn MW. **2014**. Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLOS Computational Biology* **10**: e1003998.

Doležel J, Bartoš J. **2005**. Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of botany* **95**: 99–110.

Duan H, Jones AW, Hewitt T, Mackenzie A, Hu Y, Sharp A, Lewis D, Mago R, Upadhyaya NM, Rathjen JP, *et al.* 2022. Physical separation of haplotypes in dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies with Hi-C data. *Genome Biology* **23**: 1–27.

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

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, *et al.* 2019. The Pfam protein families database in 2019. *Nucleic Acids Research* **47**: D427–D432.

Evelin H, Kapoor R, Giri B. **2009**. Arbuscular mycorrhizal fungi in alleviation of salt stress: a review. *Annals of Botany* **104**: 1263–1280.

Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma BPHJ. **2016**. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research* **26**: 1091–1100.

Falk M, Feodorova Y, Naumova N, Imakaev M, Lajoie BR, Leonhardt H, Joffe B, Dekker J, Fudenberg G, Solovei I, *et al.* 2019. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. **2020**. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America* **117**: 9451–9457.

Fortin J-P, Hansen KD. **2015**. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biology 2015 16:1* **16**: 1–23.

Galazka JM, Klocko AD, Uesaka M, Honda S, Selker EU, Freitag M. **2016**. Neurospora chromosomes are organized by blocks of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. *Genome Research* **26**: 1069–1080.

Gamper HA, van der Heijden MGA, Kowalchuk GA. **2010**. Molecular trait indicators: Moving beyond phylogeny in arbuscular mycorrhizal ecology. *New Phytologist* **185**: 67–82.

Garmaroodi HS, Taga M. **2007**. Duplication of a conditionally dispensable chromosome carrying pea pathogenicity (PEP) gene clusters in Nectria haematococca. *Molecular Plant-Microbe Interactions* **20**: 1495–1504.

Gehrmann T, Pelkmans JF, Ohm RA, Vos AM, Sonnenberg ASM, Baars JJP, Wösten HAB, Reinders MJT, Abeel T. **2018**. Nucleus-specific expression in the multinuclear mushroom-forming fungus Agaricus bisporus reveals different nuclear regulatory programs. *Proceedings of the National Academy of Sciences of the United States of America* **115**: 4429–4434.

Gel B, Serra E. **2017**. KaryoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**: 3088–3090.

Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. **2017**. Scaffolding of long read assemblies using long range contact information. *BMC genomics* **18**.

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. **2019**. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Computational Biology* **15**: e1007273.

Gianinazzi S, Vosátka M. **2011**. Inoculum of arbuscular mycorrhizal fungi for production systems: science meets business. *https://doi.org/10.1139/b04-072* **82**: 1264–1271.

Giovannetti M, Sbrana C, Logi C. **2000**. Microchambers and video-enhanced light microscopy for monitoring cellular events in living hyphae of arbuscular mycorrhizal fungi. *Plant and Soil 2000 226:2* **226**: 153–159.

Goodwin SB, M'Barek S ben, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, van der Lee TAJ, Grimwood J, Aerts A, *et al.* 2011. Finished Genome of the Fungal Wheat Pathogen Mycosphaerella graminicola Reveals Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis. *PLOS Genetics* **7**: e1002070.

Gordo I, Charlesworth B. **2000**. The degeneration of asexual haploid populations and the speed of Muller's ratchet. *Genetics* **154**: 1379–1387.

Greilhuber J, Doležel J, Lysák MA, Bennett MD. **2005**. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. *Annals of botany* **95**: 255–260.

Grigoriev I v., Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA, *et al.* 2012. The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Research* **40**: D26–D32.

Halary S, Malik SB, Lildhar L, Slamovits CH, Hijri M, Corradi N. **2011**. Conserved meiotic machinery in Glomus spp., a putatively ancient asexual fungal lineage. *Genome biology and evolution* **3**: 950–958.

Hancock JM, Bishop MJ. **2004**. HMMer. In: Dictionary of Bioinformatics and Computational Biology.

van der Heijden MGA, Klironomos JN, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller T, Wiemken A, Sanders IR. **1998**. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature 1998 396:6706* **396**: 69–72.

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

Heitman J, Kronstad JW, Taylor JW, Casselton LA. **2007**. Sex in fungi: molecular determination and evolutionary implications. *Sex in fungi: molecular determination and evolutionary implications.*

Hijri M, Sanders IR. **2005**. Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei. *Nature 2004 433:7022* **433**: 160–163.

Hoeksema JD, Chaudhary VB, Gehring CA, Johnson NC, Karst J, Koide RT, Pringle A, Zabinski C, Bever JD, Moore JC, *et al.* 2010. A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. *Ecology Letters* **13**: 394–407.

Jany J, Pawlowska TE. **2010**. Multinucleate Spores Contribute to Evolutionary Longevity of Asexual Glomeromycota. *The American Naturalist* **175**: 424–435.

Jerković I, Szabo Q, Bantignies F, Cavalli G. **2020**. Higher-Order Chromosomal Structures Mediate Genome Function. *Journal of Molecular Biology* **432**: 676–681.

Jinks JL. **1952**. Heterokaryosis; a system of adaption in wild fungi. *Proceedings of the Royal Society of London. Series B, Biological sciences* **140**: 83–99.

Judson OP, Normark BB. **1996**. Ancient asexual scandals. *Trends in Ecology & Evolution* **11**: 41–46.

Kamel L, Keller-Pearson M, Roux C, Ané JM. **2017a**. Biology and evolution of arbuscular mycorrhizal symbiosis in the light of genomics. *New Phytologist* **213**: 531–536.

Kamel L, Tang N, Malbreil M, San Clemente H, Le Marquer M, Roux C, dit Frey NF. **2017b**. The comparison of expressed candidate secreted proteins from two arbuscular mycorrhizal fungi unravels common and specific molecular tools to invade different host plants. *Frontiers in Plant Science* **8**.

Kempfer R, Pombo A. **2020**. Methods for mapping 3D chromosome architecture. *Nature Reviews Genetics* **21**: 207–226.

Kim S, Liachko I, Brickner DG, Cook K, Noble WS, Brickner JH, Shendure J, Dunham MJ. **2017**. The dynamic three-dimensional organization of the diploid yeast genome. *eLife* **6**.

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

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

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

Kokkoris V, Chagnon PL, Yildirir G, Clarke K, Goh D, MacLean AM, Dettman J, Stefani F, Corradi N. **2021**. Host identity influences nuclear dynamics in arbuscular mycorrhizal fungi. *Current Biology* **31**: 1531-1538.e6.

Kokkoris V, Stefani F, Dalpé Y, Dettman J, Corradi N. **2020**. Nuclear Dynamics in the Arbuscular Mycorrhizal Fungi. *Trends in Plant Science* **25**: 765–778.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. **2017**. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research* **27**: 722–736.

Koske RE. **1981**. Gigaspora gigantea: Observations on Spore Germination of a VA-Mycorrhizal Fungus. *Mycologia* **73**: 288.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. **2001**. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* **305**: 567–580.

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

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. **2004**. Versatile and open software for comparing large genomes. *Genome biology* **5**: 12.

Lee EH, Eo JK, Ka KH, Eom AH. **2013**. Diversity of Arbuscular Mycorrhizal Fungi and Their Roles in Ecosystems. *Mycobiology* **41**: 121.

Li H. **2013**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R. **2015**. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research* **43**: W580–W584.

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

Li X, Huang X, Chen G, Zou L, Wei L, Hua J. **2018**. Complete genome sequence of the sesame pathogen Ralstonia solanacearum strain SEPPX 05. *Genes and Genomics* **40**: 657–668.

Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, Raley C, Miller ME, Silverstein KAT, Henningsen E, *et al.* 2019. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. *Nature Communications 2019 10:1* **10**: 1–15.

Limpens E, van Creij J, Auxier B, An J, Wijfjes R, Bergin C, Rosling A, Bisseling T, Pan Z. **2022**. Stochastic nuclear organization and host-dependent allele contribution in Rhizophagus irregularis. *Research Square*.

Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DGO, Mu D, Pang E, Cao H, Cha H, Lin T, *et al.* 2014. Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an Endomycorrhizal Fungus. *PLoS Genetics* **10**: 1004078.

Logi C, Sbrana C, Giovannetti M. **1998**. Cellular events involved in survival of individual arbuscular mycorrhizal symbionts growing in the absence of the host. *Applied and Environmental Microbiology* **64**: 3473–3479.

Loureiro J, Rodriguez E, Doležel J, Santos C. **2007**. Two New Nuclear Isolation Buffers for Plant DNA Flow Cytometry: A Test with 37 Species. *Annals of Botany* **100**: 875.

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

Luginbuehl LH, Menard GN, Kurup S, van Erp H, Radhakrishnan G v., Breakspear A, Oldroyd GED, Eastmond PJ. **2017**. Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. *Science* **356**: 1175–1178.

Lutzoni F, Nowak MD, Alfaro ME, Reeb V, Miadlikowska J, Krug M, Arnold AE, Lewis LA, Swofford DL, Hibbett D, *et al.* 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications* **9**: 1–11.

Maeda T, Kobayashi Y, Kameoka H, Okuma N, Takeda N, Yamaguchi K, Bino T, Shigenobu S, Kawaguchi M. **2018a**. Evidence of non-tandemly repeated rDNAs and their intragenomic heterogeneity in Rhizophagus irregularis. *Communications Biology* **1**.

Maeda T, Kobayashi Y, Kameoka H, Okuma N, Takeda N, Yamaguchi K, Bino T, Shigenobu S, Kawaguchi M. **2018b**. Evidence of non-tandemly repeated rDNAs and their intragenomic heterogeneity in Rhizophagus irregularis. *Communications Biology 2018 1:1* **1**: 1–13.

Malar C M, Krüger M, Krüger C, Wang Y, Stajich JE, Keller J, Chen ECH, Yildirir G, Villeneuve-Laroche M, Roux C, *et al.* 2021. The genome of *Geosiphon pyriformis* reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Current Biology* **31**: 1570-1577.e4.

Malar C M, Wang Y, Stajich JE, Kokkoris V, Villeneuve-Laroche M, Yildirir G, Corradi N. **2022**. Early branching arbuscular mycorrhizal fungus *Paraglomus occultum* carries a small and repeat-poor genome compared to relatives in the Glomeromycotina. *Microbial Genomics* **8**: 000810.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. **2018**. MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology* **14**.

Marroni F, Pinosio S, Morgante M. **2014**. Structural variation and genome complexity: is dispensable really dispensable? *Current Opinion in Plant Biology* **18**: 31–36.

Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, *et al.* 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*.

Martin FM, Uroz S, Barker DG. **2017**. Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356**.

Mathieu S, Cusant L, Roux C, Corradi N. **2018**. Arbuscular mycorrhizal fungi: intraspecific diversity and pangenomes. *New Phytologist* **220**: 1129–1134.

Matthey-Doret C, Colp MJ, Escoll P, Thierry A, Moreau P, Curtis B, Sahr T, Sarrasin M, Gray MW, Lang BF*, et al.* 2022. Chromosome-scale assemblies of *Acanthamoeba castellanii* genomes provide insights into *Legionella pneumophila* infection–related chromatin reorganization. *Genome Research*.

McCarthy CGP, Fitzpatrick DA. **2019**. Pan-genome analyses of model fungal species. *Microbial Genomics* **5**.

McDonald MC, Taranto AP, Hill E, Schwessinger B, Liu Z, Simpfendorfer S, Milgate A, Solomon PS. **2019**. Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. *mBio* **10**.

Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. **2005**. The microbial pan-genome. *Current Opinion in Genetics & Development* **15**: 589–594.

Meraldi P, McAinsh AD, Rheinbay E, Sorger PK. **2006**. Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biology* **7**: 1–21.

Montoliu-Nerin M, Sánchez-García M, Bergin C, Grabherr M, Ellis B, Kutschera VE, Kierczak M, Johannesson H, Rosling A. **2020**. Building de novo reference genome assemblies of complex eukaryotic microorganisms from single nuclei.

Morin E, Miyauchi S, San Clemente H, Chen ECH, Pelin A, de la Providencia I, Ndikumana S, Beaudet D, Hainaut M, Drula E*, et al.* 2019. Comparative genomics of Rhizophagus irregularis, R. cerebriforme, R. diaphanus and Gigaspora rosea highlights specific genetic features in Glomeromycotina. *New Phytologist* **222**: 1584–1598.

Naranjo-Ortiz MA, Gabaldón T. **2019**. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biological Reviews* **94**: 2101–2137.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. **2015**. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Normark BB, Judson OP, Moran NA. **2003**. Genomic signatures of ancient asexual lineages. *Biological Journal of the Linnean Society* **79**: 69–84.

Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. **2016**. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology* **17**: 1–14.

Öpik M, Davison J. **2016**. Uniting species- and community-oriented approaches to understand arbuscular mycorrhizal fungal diversity. *Fungal Ecology* **24**: 106–113.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. **2017**. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**: 417–419.

Pawlowska TE, Taylor JW. **2004**. Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature 2004 427:6976* **427**: 733–737.

Pawlowska TE, Taylor JW. **2005**. Hyphal fusion and multigenomic structure (reply). *Nature 2005 433:7022* **433**: E4–E4.

Petersen TN, Brunak S, von Heijne G, Nielsen H. **2011**. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785–786.

Pires AS, Azinheira HG, Cabral A, Tavares S, Tavares D, Castro M, Várzea V, Silva MC, Abranches R, Loureiro J, *et al.* 2016. Cytogenomic characterization of *Colletotrichum kahawae* , the causal agent of coffee berry disease, reveals diversity in minichromosome profiles and genome size expansion. *Plant Pathology* **65**: 968–977.

Pirozynski KA, Malloch DW. **1975**. The origin of land plants: A matter of mycotrophism. *BioSystems* **6**: 153–164.

Plett JM, Martin F. **2012**. Poplar root exudates contain compounds that induce the expression of MiSSP7 in Laccaria bicolor. *Plant Signaling and Behavior* **7**.

Plissonneau C, Daverdin G, Ollivier B, Blaise F, Degrave A, Fudal I, Rouxel T, Balesdent MH. **2016**. A game of hide and seek between avirulence genes AvrLm4-7 and AvrLm3 in Leptosphaeria maculans. *New Phytologist* **209**: 1613–1624.

Pombo A, Dillon N. **2015**. Three-dimensional genome architecture: Players and mechanisms. *Nature Reviews Molecular Cell Biology* **16**: 245–257.

Pontecorvo G, Roper JA, Forbes E. **1953**. Genetic recombination without sexual reproduction in Aspergillus niger. *Journal of general microbiology* **8**: 198–210.

Quinlan AR, Hall IM. **2010**. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. **2018**. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications* **9**.

Rayner ADM. **1991**. The Challenge of the Individualistic Mycelium. *Mycologia* **83**: 48.

Redecker D, Kodner R, Graham LE. **2000**. Glomalean fungi from the Ordovician. *Science* **289**: 1920–1921.

Redecker D, Raab P. **2017**. Phylogeny of the Glomeromycota (arbuscular mycorrhizal fungi): recent developments and new gene markers View supplementary material Phylogeny of the Glomeromycota (arbuscular mycorrhizal fungi): recent developments and new gene markers.

Reinhardt D, Roux C, Corradi N, Di Pietro A. **2021**. Lineage-Specific Genes and Cryptic Sex: Parallels and Differences between Arbuscular Mycorrhizal Fungi and Fungal Pathogens. *Trends in Plant Science* **26**: 111–123.

Remy W, Taylor TN, Hass H, Kerp H. **1994**. Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 11841–11843.

Rich MK, Nouri E, Courty PE, Reinhardt D. **2017**. Diet of Arbuscular Mycorrhizal Fungi: Bread and Butter? *Trends in Plant Science* **22**: 652–660.

Riley R, Charron P, Idnurm A, Farinelli L, Dalpé Y, Martin F, Corradi N. **2014**. Extreme diversification of the mating type-high-mobility group (MATA-HMG) gene family in a plant-associated arbuscular mycorrhizal fungus. *New Phytologist* **201**: 254–268.

Roach MJ, Schmidt SA, Borneman AR. **2018**. Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**: 1–10.

Rodríguez-Echeverría S, Teixeira H, Correia M, Timóteo S, Heleno R, Öpik M, Moora M. **2017**. Arbuscular mycorrhizal fungi communities from tropical Africa reveal strong ecological structure. *New Phytologist* **213**: 380–390.

Ropars J, Toro KS, Noel J, Pelin A, Charron P, Farinelli L, Marton T, Krüger M, Fuchs J, Brachmann A, *et al.* 2016. Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nature Microbiology* **1**: 16033.

Ryan FJ, Lederberg J. **1946**. Reverse-Mutation and Adaptation in Leucineless Neurospora. *Proceedings of the National Academy of Sciences of the United States of America* **32**: 163.

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

Sanders IR. **2002**. Ecology and Evolution of Multigenomic Arbuscular Mycorrhizal Fungi. *The American Naturalist* **160**: S128–S141.

Sanders IR, Croll D. **2010**. Arbuscular Mycorrhiza: The Challenge to Understand the Genetics of the Fungal Partner. *Annual Review of Genetics* **44**: 271–292.

Sanyal K, Baum M, Carbon J. **2004**. Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 11374–11379.

Savary R, Masclaux FG, Wyss T, Droh G, Cruz Corella J, Machado AP, Morton JB, Sanders IR. **2017**. A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus Rhizophagus irregularis. *The ISME Journal 2018 12:1* **12**: 17–30.

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

Schmieder R, Lim YW, Rohwer F, Edwards R. **2010**. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* **11**: 1–14.

Schmitt AD, Hu M, Ren B. **2016**. Genome-wide mapping and analysis of chromosome architecture. *Nature Reviews Molecular Cell Biology* **17**: 743–755.

Schurko AM, Neiman M, Logsdon JM. **2009**. Signs of sex: what we know and how we know it. *Trends in Ecology & Evolution* **24**: 208–217.

Schüßler A, Schwarzott D, Walker C. **2001**. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research* **105**: 1413–1421.

Schwessinger B. **2016**. High quality DNA from Fungi for long read sequencing e.g. Pacbio. *Protocols.io*: 1–6.

Seidl MF, Thomma BPHJ. **2014**. Sex or no sex: Evolutionary adaptation occurs regardless. *BioEssays* **36**: 335–345.

Selmecki A, Forche A, Berman J. **2010**. Genomic plasticity of the human fungal pathogen Candida albicans. *Eukaryotic Cell* **9**: 991–1008.

Selosse MA, le Tacon F. **1998**. The land flora: a phototroph-fungus partnership? *Trends in ecology & evolution* **13**: 15–20.

Serghi EU, Kokkoris V, Cornell C, Dettman J, Stefani F, Corradi N. **2021**. Homo- and Dikaryons of the Arbuscular Mycorrhizal Fungus Rhizophagus irregularis Differ in Life History Strategy. *Frontiers in Plant Science* **12**: 1544.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. **2015**. HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biology* **16**.

Shakya M, Ahmed SA, Davenport KW, Flynn MC, Lo CC, Chain PSG. **2020**. Standardized phylogenetic and molecular evolutionary analysis applied to species across the microbial tree of life. *Scientific Reports 2020 10:1* **10**: 1–15.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E v., Zdobnov EM. **2015**. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Smit AFA, Hubley R, Green P. **2013**. RepeatMasker Open-4.0.

Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent MH, Connolly LR, Freitag M, Rouxel T, *et al.* 2014. Epigenetic Control of Effector Gene Expression in the Plant Pathogenic Fungus Leptosphaeria maculans. *PLoS Genetics* **10**: e1004227.

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EVL, Brown JKM, Butcher SA, Gurr SJ, *et al.* 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science (New York, N.Y.)* **330**: 1543–1546.

Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A, *et al.* 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**: 1028.

Sperschneider J, Dodds PN. **2021**. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *bioRxiv*: 454080.

Sperschneider J, Jones AW, Nasim J, Xu B, Jacques S, Zhong C, Upadhyaya NM, Mago R, Hu Y, Figueroa M, *et al.* 2021. The stem rust fungus Puccinia graminis f. sp. tritici induces centromeric small RNAs during late infection that are associated with genome-wide DNA methylation. *BMC Biology 2021 19:1* **19**: 1–25.

Spielmann M, Lupiáñez DG, Mundlos S. **2018**. Structural variation in the 3D genome. *Nature Reviews Genetics* **19**: 453–467.

Springer-Verlag ©, Wagner F, Gay G, Debaud JC. **1988**. Genetical variability of glutamate dehydrogenase activity in monokaryotic and dikaryotic mycelia of the ectomycorrhizal fungus Hebeloma cylindrosporum. *Appl Microbiol Biotechnol* **28**: 566–571.

Steinkraus DC, Kramer JP. **1989**. Development of Resting Spores of Erynia aquatica (Zygomycetes: Entomophthoraceae) in Aedes aegypti (Diptera: Culicidae). *Environmental Entomology* **18**: 1147–1152.

Stockinger H, Walker C, Schüßler A. **2009**. 'Glomus intraradices DAOM197198', a model fungus in arbuscular mycorrhiza research, is not Glomus intraradices. *New Phytologist* **183**: 1176–1187.

Stukenbrock EH. **2013**. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytologist* **199**: 895–907.

Sun X, Chen W, Ivanov S, MacLean AM, Wight H, Ramaraj T, Mudge J, Harrison MJ, Fei Z. **2019**. Genome and evolution of the arbuscular mycorrhizal fungus Diversispora epigaea (formerly Glomus versiforme) and its bacterial endosymbionts. *New Phytologist* **221**: 1556–1573.

Szabo Q, Bantignies F, Cavalli G. **2019**. Principles of genome folding into topologically associating domains. *Science Advances* **5**.

Talhinhas P, Tavares D, Ramos AP, Gonçalves S, Loureiro J. **2017**. Validation of standards suitable for genome size estimation of fungi. *Journal of microbiological methods* **142**: 76–78.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, Frei Dit Frey N, Gianinazzi-Pearson V, *et al.* 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 20117–20122.

Vandenkoornhuyse P, Leyval C, Bonnin I. **2001**. High genetic diversity in arbuscular mycorrhizal fungi: evidence for recombination events. *Heredity* **87**: 243–253.

Vaser R, Sović I, Nagarajan N, Šikić M. **2017**. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* **27**: 737–746.

Vernikos G, Medini D, Riley DR, Tettelin H. **2015**. Ten years of pan-genome analyses. *Current Opinion in Microbiology* **23**: 148–154.

Vlaardingerbroek I, Beerens B, Schmidt SM, Cornelissen BJC, Rep M. **2016**. Dispensable chromosomes in Fusarium oxysporum f. Sp. lycopersici. *Molecular Plant Pathology* **17**: 1455–1466.

Vosátka M, Látr A, Gianinazzi S, Albrechtová J. Development of arbuscular mycorrhizal biotechnology and industry: current achievements and bottlenecks.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**: 112963.

Walker C, Harper CJ, Brundrett MC, Krings M. **2018**. Looking for Arbuscular Mycorrhizal Fungi in the Fossil Record. *Transformative Paleobotany*: 481–517.

Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, *et al.* 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research* **42**: D581.

Welch DM, Meselson M. **2000**. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* **288**: 1211–1215.

Winter DJ, Ganley ARD, Young CA, Liachko I, Schardl CL, Dupont PY, Berry D, Ram A, Scott B, Cox MP. **2018**. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus Epichloë festucae. *PLoS Genetics* **14**.

Wolff J, Bhardwaj V, Nothjunge S, Richard G, Renschler G, Gilsbach R, Manke T, Backofen R, Ramírez F, Grüning BA. **2018**. Galaxy HiCExplorer: A web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Research* **46**: W11–W16.

Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, Grüning BA. **2020**. Galaxy HiCExplorer 3: A web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Research* **48**: W177–W184.

Wyss T, Masclaux FG, Rosikiewicz P, Pagni M, Sanders IR. **2016**. Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus Rhizophagus irregularis. *The ISME Journal 2016 10:10* **10**: 2514–2526.

Xu J, Horgen PA, Anderson JB. **1996**. Somatic recombination in the cultivated mushroom Agaricus bisporus. *Mycological Research* **100**: 188–192.

Yildirir G, Malar C M, Kokkoris V, Corradi N. **2020**. Parasexual and Sexual Reproduction in Arbuscular Mycorrhizal Fungi: Room for Both. *Trends in Microbiology* **28**: 517–519.

Yildirir G, Sperschneider J, Malar C M, Chen EC, Iwasaki W, Cornell C, Corradi N. **2021**. Long reads and Hi-C sequencing illuminate the two compartment genome of the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytologist*.

Zeng T, Holmer R, Hontelez J, te Lintel-Hekkert B, Marufu L, de Zeeuw T, Wu F, Schijlen E, Bisseling T, Limpens E. **2018**. Host- and stage-dependent secretome of the arbuscular mycorrhizal fungus Rhizophagus irregularis. *Plant Journal* **94**: 411–425.

Zheng H, Xie W. **2019**. The role of 3D genome organization in development and cell differentiation. *Nature Reviews Molecular Cell Biology* **20**: 535–550.