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, Australia

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

* contributed equally

Corresponding Author

Abstract

Chromosome folding links genome structure with gene function by generating distinct nuclear compartments and topologically associating domains (TADs). In mammals, these domains undergo preferential interactions and regulate gene expression, however in fungi the role of chromosome folding in genome biology is unclear. Here, we combine Nanopore (ONT) sequencing with chromatin conformation capture sequencing (Hi-C) to reveal chromosome diversity in a group of obligate plant symbionts with a multinucleate mycelium; the arbuscular mycorrhizal fungi (AMF). We find that phylogenetically distinct strains of the model AMF *Rhizophagus irregularis* all carry 33 chromosomes. Homologous chromosomes show within species variability in size, as well as in gene and repeat content. Strain-specific Hi-C sequencing reveals that all strains have a 3D genome organization that resembles a 'checkerboard' structure with two distinct (A/B) chromatin compartments. Each compartment differs in the level of gene transcription, regulation of candidate effectors and methylation rate. 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 mechanisms that generate and modify chromosome folding during host-microbe interactions.

Introduction

Arbuscular mycorrhizal fungi (AMF) are obligate plant mutualistic organisms of the fungal subphylum Glomeromycotina (1) that evolved a symbiotic relationship with the roots of most vascular plants species (2–4). During the AM symbiosis, these fungi enhance the plant host nutrient and water uptake in exchange of sugars and lipids, and can also provide their host with increased resistance against pathogens (5). 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 across all continents, highlighting AMF's capacity to persist in various environmental conditions (6). AMF spores and hyphae are peculiar as they harbor thousands of coexisting nuclei at all times (7). The genetics of these fungi are defined by the presence of two life stages – i.e. a homokaryotic stage (AMF homokaryons) where co-existing nuclei are genetically largely uniform, and an heterokaryotic stage (AMF dikaryons) where thousands of nuclei originating from two parental strains coexist in one large cell, presumably as a result of sexual reproduction (8–11).

To date, existing genome resources have revealed important information about the biology and ecology of Glomeromycotina. In particular, all species within this subphylum carry gene losses in cellular pathways that produce fatty acids and sugars as well as plant cell wall degradation, highlighting their host dependence for nutrition (12–16). Comparative genomic analyses revealed that AMF carry an abundance of transposable elements (TE) and more recently, that closely related AMF strains vary dramatically in their gene content (14, 17).

The presence of high genome diversity within species indicated that AMF have pangenomes (14, 17). The pangenome concept highlights the distinction between 'core' conserved genes and 'dispensable ' non-conserved genes among strains of a species. Within pangenomes, core genes are hypothesized to be essential for daily activities (e.g. cytoskeleton, protein translation etc.), while dispensable genes are thought to help individual strains adapt to changing environments and/or to associate with distinct hosts (e.g. transduction signalling, effectors, virulence genes, etc.) (18–20).

Identifying the mechanisms from which within-species genome diversity emerges is essential for understanding the mode of evolution of these widespread plant symbionts. In fungal plant

pathogens, high repeat content was found to generate within species genome variability, and it was consequently proposed that similar mechanisms could also drive genetic variability in AMF (14, 17). For AMF, however this hypothesis is currently untestable because most available genome assemblies (including most of those obtained from the model *R. irregularis* strains) are highly fragmented – i.e. composed in many cases of dozens of thousands of scaffolds or contigs (14, 21). This fragmentation generates mosaic sequences of multiple haplotypes (collapsed regions), chimeric assemblies, and artificial gene duplications (22, 23), overall 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 has also hampered the identification of mechanisms that determine the relationship between genome structure and function. Within nuclei, chromosomes can be separated into either active open chromatin A-compartments or inactive closed chromatin B-compartments, and genes within each compartment can be differently regulated (24–27). This distinct, compartment-driven gene regulation highlights the epigenetic role of chromosome folding for mammalian biology, but in fungi knowledge of this process was thus far restricted to a few model genera (e.g. *Sacharomyces*, *Neurospora*) and the filamentous fungus *Epichloë festucae* (28, 29). In the plant symbiont *E. festucae* chromosome folding and repetitive elements together divide the genome into distinct regions with very similar gene expression profiles, suggesting that both genome features co-regulate the symbiotic associations formed by this species with different plant hosts. However, whether similar mechanisms also drive gene expression and symbiotic relationships in other fungal symbionts, including AMF, is unknown.

Here, we aimed to improve our understanding of within species genomic and epigenetic variability in the model mycorrhizal symbiont *R. irregularis* by combining long-reads sequencing and strain-specific high-throughput chromatin conformation capture (Hi-C) (30). This approach produced complete chromosome assemblies and high-resolution epigenetic information - i.e. methylation and chromosome folding - from five phylogenetically distinct homokaryotic strains of the species *R. irregularis* (DAOM-197198, A1, C2, B3, 4401) with previously fragmented genomes.

Materials and Methods

Culturing, DNA Extraction and ONT/Hi-C Sequencing

R. irregularis strains A1 (DAOM-664342), B3 (DAOM-664345) and C2 (DAOM-664346) were originally isolated from Switzerland (Tänikon), while the strain DAOM-197198 was originally isolated from Pont Rouge, Canada (31, 32) and the strain 4401 (DAOM-240446) was isolated from *Ammophila breviligulata* in "La Martinique Îles-de-la-Madeleine" (Canada).

All five strains were cultured *in vivo*, using *Daucus carota* root organ cultures (ROCs) as the host, as previously described (33). 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 (34). Following extraction, high quality DNA samples were processed using Nanopore Ligation Sequencing Kit SQK-LSK109 to prepare the sequencing libraries, which we sequenced using the MinION R9.4.1 flowcells to produce an average of 8.45 million reads with an N50 > than 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 used to generate assemblies using Canu (version 2.0) (35). These assemblies were then further polished using Racon (v 1.4.10) (36) using the default parameters and consensus sequences were created using Medaka (version 1.0.3). The polishing step was concluded by two rounds of Pilon (version 1.23) using Illumina reads and default parameters (37). Illumina reads were mapped on the polished contigs using bwa mem (version 0.7.17) (38) and the read coverage of each contig was calculated using bedtools (version 2.26.0) (39). Contigs having abnormally low (<5x) and high (>200x) coverage were filtered out using purge_haplotigs (40), 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*. (8), DAOM197198 reads were obtained from Chen et al. 2018(14), 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 by 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 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 (38). Alignments were then processed with the Arima Genomics pipeline (41). Scaffolding was performed using SALSA 2.2 (42, 43) and subsequent manual curation was guided by Hi-C contact maps. For DAOM197198, the final genome assembly of Maeda et al. (44) was scaffolded further with SALSA 2.2, also using strain-specific Hi-C data.

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; (14)). 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) (45).

Conserved protein domains were predicted using Pfam v.27 (46). SignalP 4.1 (-t euk -u 0.34 -U 0.34)(47) and TMHMM 2.0(48) were used to predict secreted proteins. A protein was called secreted if it was predicted to have a signal peptide and has no transmembrane domains. Effector candidates were predicted with EffectorP 3.0(49). De novo repeats were predicted with RepeatModeler 2.0.0 and the option -LTRStruct (50). These were merged with the RepeatMasker repeat library and RepeatMasker 4.1.0 was run with this combined repeat database (51). 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 (52, 53) 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) (54) for *k-mer* distance calculations and dnadiff from the MUMmer (55, 56) suite for structural variation calculations between strains and compartments. Orthology analyses were made by FastOrtho (57), using following parameters: 50% identity and 50% coverage using protein sequences of all five assemblies. All karyoplots were produced using KaryoploteR (58). All genome data and reads newly obtained are available in Genbank under the BioProject PRJNA748024 (pending release by NCBI).

Identification of chromosome compartments and topologically associated domains

We called A/B compartments and TADs with hicexplorer 3.6 (59) using the commands hicPCA and hicFindTADs, respectively. Inter- and intra-chromosomal Hi-C contact maps were produced with HiC-Pro 2.11.1 (60) (MAPQ=10) and hicexplorer 3.6 (28, 59, 61, 62) and these Hi-C contact maps were manually inspected to assign chromosomal regions to A/B compartments. Briefly, the regions along the same chromosome carrying the same PCA1 values – i.e. positive or negative - belong to the same compartment. Regions found to be in close physical proximity to the smallest chromosomes being are assigned to compartment A, while those carrying the counterpart PCA1 signal were assigned to compartment B. This process was repeated for all 33 chromosomes individually to separate the chromosomal regions into separate compartments. Bedtools (39) 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 (63) 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 (64) to assess gene differential expression (padj < 0.1). The DAOM-197198 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 (35), and then scaffolded using strain-specific Hi-C data. The same scaffolding procedures were carried out on a previously published, polished PacBio assembly of the model AMF strain DAOM197198 (44). 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 (**Figure 1, Table 1, Figure S1**).

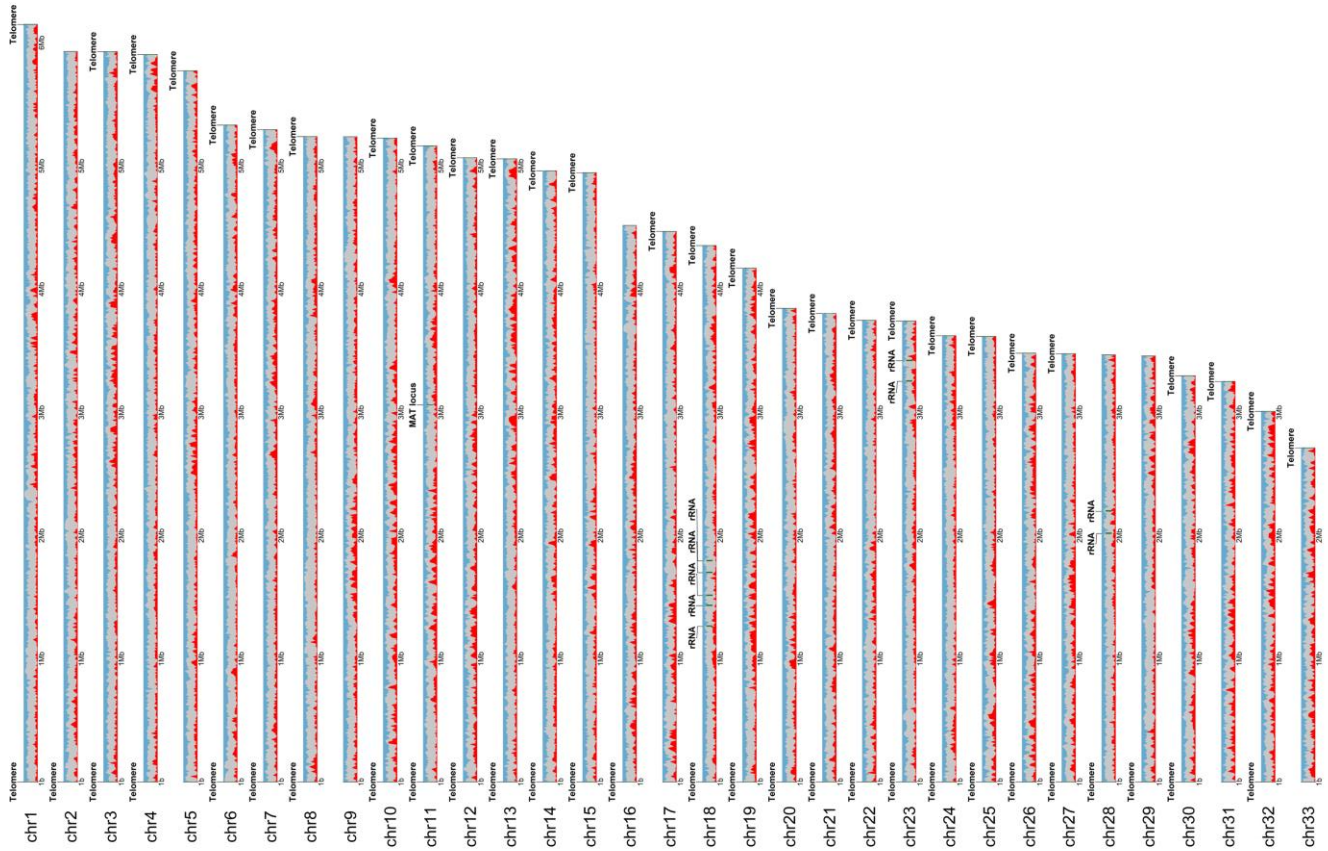

Figure 1: R. irregularis DAOM-197198 strain chromosomes. The length of all 33 chromosomes is indicated in the karyoplot. Telomere repeats, MAT locus and small subunit rRNA locations are marked. Gene densities are shown in red, and repeat densities are shown in blue.

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 DAOM-197198, 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 (8). 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 (L50 26-51 Kb), indicating that the chromosome assemblies are near-complete.

Table 1: Statistics for chromosome-scale assemblies for the five strains.

	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

We identified an average of 76 families of repeats encompassing on average 50.2% of the genome, and these include known transposable elements (TEs) (**Figure S2**). When only TE's are considered, repeat density varies across chromosomes of different strains. For DAOM-197198, 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 DAOM-197198 (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 (14), and BUSCO completeness averages 96% (**Table 1**). Gene density also varies among chromosomes, with chromosome 32 always having the highest density, while

chromosome 15 has the lowest (**Figure 1, Figures S1**). A characteristic of AMF is that they carry divergent rRNA operons within their genome. In DAOM0197198, two rRNA operons co-locate on chromosomes 23 and 28, and five are present in chromosome 18 (**Figure 1**). Among other notable AMF genes, those involved in plant cell wall degradation (12–16) are all located on distinct chromosomes, while the putative MAT-locus (8) is located on chromosome 11 in all strains.

Within species chromosome diversity in R. irregularis

Consistent with previous reports on fragmented assemblies, all strains show substantial variability in gene numbers and putative functions (14) . The *R. irregularis* gene content is divided into genes shared by all strains (core genes) and dispensable genes shared by only a few strains (or strain-specific) (**Figure S3a**), and each strain carries a distinct abundance of protein domains (**Figure S3b**). The average ratio between core and dispensable genes is almost identical from those reported using fragmented assemblies – i.e. on average only 55.8% of *R. irregularis* genes are core (14), and in addition to protein encoding genes our analyses confirm that within species variability also affects the number of rRNA operons (**Table S1**) (65). Homologous chromosomes also 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. Similarly, 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.

We used a *k*-mer distance measure to assess relatedness of strains and their chromosomes. The model strain DAOM197198 is closely related to strain 4401, strains B3 and A1 are closely related and C2 is most distant to the other strains (**Figure S4**). 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 S4**), consistent with past reports of phylogenetic incongruence and putative inter-strain genetic exchange in *R. irregularis* (14, 66)**.**

The chromosomes of Rhizophagus irregularis form a two-compartment genome

Hi-C sequencing also generated direct, quantitative evidence of the nuclear organization for the five *R. irregularis* strains and revealed that all chromosomes form a distinct 'checkerboard' structure (**Figure 2**). For the strains DAOM-197198, A1 and C2, high quality Hi-C data also allowed us to confidently define the coordinates of the long-range interactions that form two

nuclear A/B compartments and the topologically-associated domains (TADs) within them (67, 68) **Figures S5-10**).

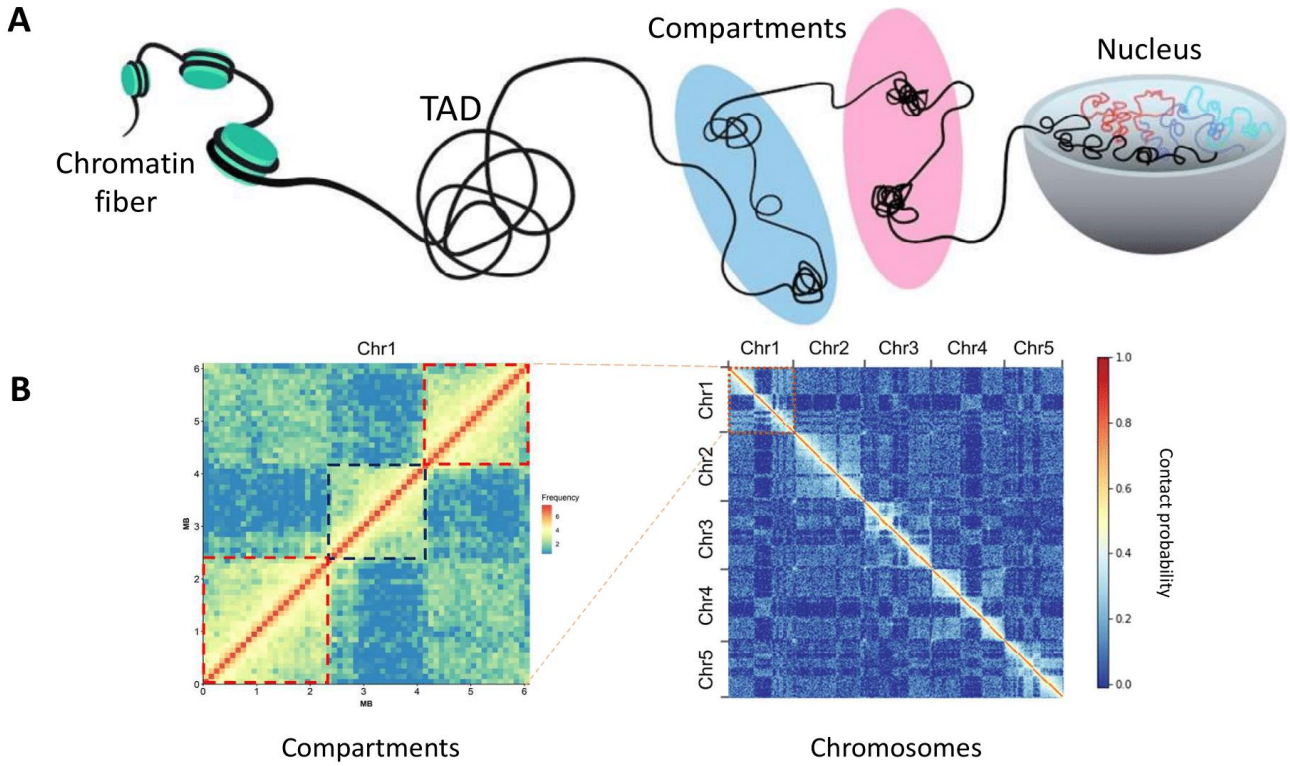

Figure 2: Chromatin folding inside of the R. irregularis nucleus. A) Schematic showing the compaction of chromatin fibers inside of an R. irregularis nucleus. Chromatin fibers physically interact with each other and may fold into regions called compartments and topologically associated domains. B) Hi-C contact maps showing the compartmentalization of the R. irregularis genome. In these heat maps created to demonstrate contact probabilities 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. The interactions of these bright squares created by increased contact frequency are then further analyzed to group them into one of the two compartments (For chr1, the compartment shown in the black square belongs to compartment A, and the other two compartment shown in red square belongs to compartment B . When the interactions of several chromosomes are analyzed, a "checkered" pattern indicates the two-compartment nature of the genome arrangement.

In the model strain DAOM-197198 the A-compartment has a total size of 58.1 Mb, while the B-compartment is 73.5 Mb, and respectively they contain 96 and 83 TADs. Each chromosome varies in the degree of compartmentalization – e.g. chromosome 2 has consistent swaps between A- and B-compartments, while the A-compartment dominates in the smallest chromosomes (chr30 – 33) **(Figure S6-10)**. In addition to compartments, Hi-C maps revealed other significant physical interactions both within and among chromosomes, including telomere-to-telomere interactions between chromosomes. However, we found that the *Rhizophagus* Hi-C contact maps did not show the easily recognizable centromere hotspots found in other fungi (69). Furthermore, we visually identified distinct changes in Hi-C contacts among strains, and inspection of sequence data

revealed that these are linked with the emergence of strain-specific repeat and gene expansions (**Figure S11a**), or inversions (**Figure S11b,c,d**).

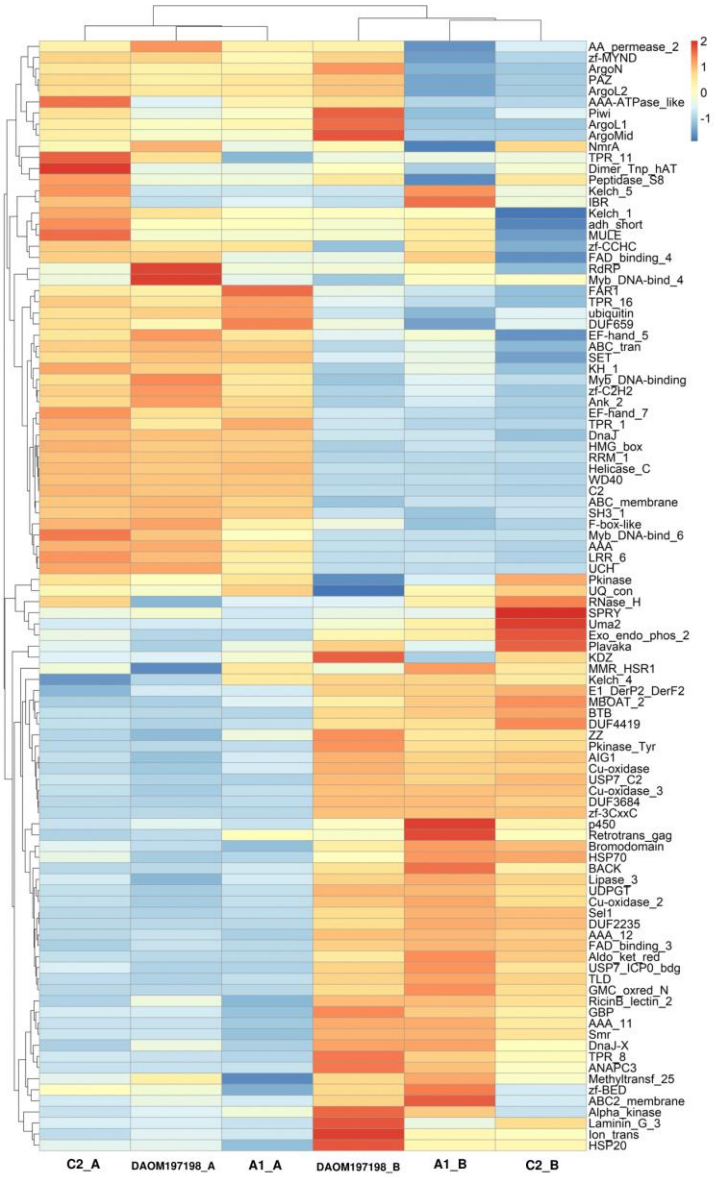

Figure 3: Pfam domains in the A/B compartments. Hierarchical clustering of t-values for PFAM domains in the A and B compartments. The pfam_scan.pl (52, 53) script was used on the annotated proteins of five assemblies and protein domains were obtained. After 100 most abundant Pfam protein domains were selected, the frequency values were transformed using a t-test. Relative enrichment (red) and relative depletion (blue) were shown. Left: A-compartment in C2, DAOM197198 and A1. Right: B-*compartment in DAOM-197198, A1 and C2*

Each compartment carries distinct gene and repeat densities. In the model strain DAOM197198, the A-compartment is more gene dense with respectively 25.6 genes and 207.3 repeats per 10 Kb, while the B-compartment carries an average of 19.5 genes and 237.7 repeats per 10 Kb (**Table S2)**. The A-compartment codes for 90% of BUSCO proteins, and double the number of core genes

(**Figure S12**), while the B-compartment carries most dispensable genes and most families known to be highly expanded in AMF – i.e. the tetratricopeptide repeat Sel-1, the homodimerization BTB, or Tyrosine Kinases. The diverse high mobility box (HMG) gene family is, however, predominant in compartment B (**Figure 3**). 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 (70).

Gene expression in chromosomal compartments and TAD regions

In other organisms, A-compartments are generally transcriptionally active while B-compartments are enriched in repressive heterochromatin (24). To test if this dichotomy also applies to AMF, we investigated gene expression between compartments in DAOM-197108 using three RNA-seq high quality datasets obtained, respectively, from intra-radical hyphae in symbiosis with *Allium*, arbuscocytes in symbiosis with *Medicago* and germinating spores (15, 71).

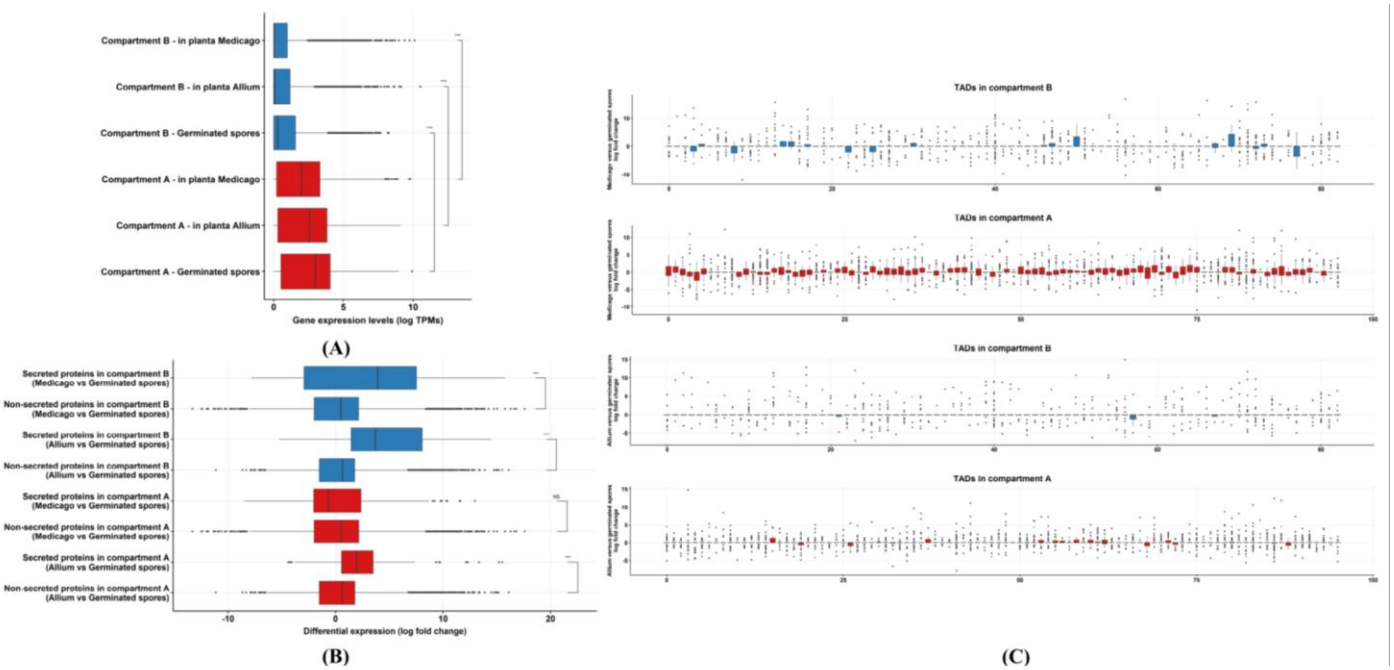

Figure 4: Gene expression analysis for the A/B compartments. (A) Genes in the A-compartment show significantly higher expression levels measured in Transcripts Per Million (TPM) than genes in the B-compartment in all three conditions (germinated spores a in germinated spores. (B) Genes encoding secreted proteins are up-regulated in both in planta conditions in the B-compartment, but not in the A-compartment. (C) Each TAD is plotted on the x-axis, with the log fold change differences (in planta versus germinated spores) for all genes within a TAD plotted on the y-axis. TADs in the A-compartment appear to be more co-regulated, particularly for the Medicago in planta RNAseq data, than TADs in the B-compartment.

Consistent with current knowledge of chromosomal compartments (72, 73), *R. irregularis* genes in the A-compartment have, on average, significantly higher expression levels than those in the B-compartment (**Figure 4a**). However, we also find that the A-compartments have 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). Additionally, we observed that genes encoding secreted proteins are significantly more up-regulated in both *in planta* conditions in the B-compartment, but not in the A-compartment (**Figure 4b**). As each compartment contains regions TAD, we also aimed to investigate the expression of these regions using the same RNA-seq datasets. This analysis revealed 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 4c)**.

Methylation in chromosomal compartments and chromosomal rearrangement

RNA-seq data obtained from three distinct conditions showed that the A-compartment produces well over three times more transcripts per million than the B-compartment. We hypothesized that this difference resulted from different methylation patterns affecting A/B compartments (74, 75). To test this, we used ONT sequencing to call methylated CpG sites and found significantly more methylation in the A-compartment, particularly for highly methylated sites (methylation frequency > 80%) (**Figure S13a)**.

Within the B-compartment, genes are more methylated than in the A-compartment. Specifically, 22.5% of all protein encoding genes are highly methylated (methylation frequency > 80%) in the B-compartment while in the A-compartment 18.4% of the gene space is highly methylated. When TEs are considered, the methylation rates are inversed. In the B-compartment of DAOM-197198 assembly, 11 out of 36 TE families (30.5%) are unmethylated compared to 5.7% (2 out of 35) in the A-compartment. Accordingly, in the A-compartment (**Figure S13b**) 67% of methylated TEs families (24 out of 35) are highly methylated compared to only 25% for the B -compartment (9 out of 36). We also found that a few TE families are compartment specific - e.g Ginger DNA 2, DNA Zisupton and LINE Penelope are only found in compartment B, while the LINE -Tad, and LINE R1-LOA carried by the A-compartment – and consistent with the methylation patterns described above, all compartment-specific TEs are exclusively methylated in the A-compartment.

We hypothesized that the differences in transcription and methylation, coupled with distinct gene and repeat densities, could result in different structural rearrangement rates between A/B

compartments. To test this, we measured the degree of structural variants and sequence identity of each compartment. These analyses showed that the sequence alignments between strains in the B-compartment are shorter and have less sequence identity than sequence alignments in the A-compartments. The B-compartment also contain more structural variation, indicating overall that each compartment follows different evolutionary trajectories driven by differences in their repeat content (**Table S3**, **Figure S13**). The strain-specific gene and repeat expansions that generate variability in Hi-C contact (**Figure S11a**) also occur in the B-compartment.

Discussion

A chromosome-level view of a model AMF pangenome

The combined number of genes identified in individuals of the model AMF *R. irregularis* was reported to exceed that found in individual strains (14, 21, 76). The present study confirms this, but also refines this view by revealing that model AMF carry 33 homologous chromosomes (possibly the largest number identified in a fungus), and that within-species variability includes chromosome size and chromosome-driven epigenetic signatures (folding). Chromosome size variation is found in all strains analyzed, regardless of their phylogenetic similarity (11, 14, 77), and even strains that carry the same putative MAT-locus – e.g. the strains C2 and 4401 both carry the MAT-locus-6 (as defined by Ropars *et al.*(8)) on chromosome 11, but this chromosome is 640 Kb larger in C2 compared to 4401. Despite this variability all strains analysed to date carry the same number of homologous chromosomes, and future studies may reveal if this trait could help define species boundaries in this taxonomically challenging fungal group (78).

Chromosome compartments dictate AMF genome biology and evolution

Hi-C data revealed that AMF chromosomes separate into locations with open chromatin (transcriptionally active A-compartment) and closed chromatin (transcriptionally less active B-compartments) that regulate both the expression and evolution of distinct functions. In the A-compartment, the higher gene density and high repeat methylation indicates a tighter control over 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 recently proposed that AMF and notorious plant pathogens (e.g. *Fusarium oxysporum*, *Verticillium dahlia*; *Zymoseptoria tritici*) evolved similar genetic strategies to cope with their evolving plant hosts and competitors (76). Within this context, 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) (17). 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 of dispensable or supernumerary chromosomes in *R. irregularis* – i.e. repeat-rich genomic regions of fungal plant pathogens that drive adaptation to changing environments and hosts by encoding virulence factors, secondary metabolite synthesis gene clusters (79–81).

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 evidence 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 (82). In the fungal pathogen *Leptosphaeria maculans* epigenetic control mechanisms lead to effector gene regulation (83), and acquiring *in planta* Hi-C data may reveal if plant colonization always lead to chromosome conformation changes that lift repression for the B-compartment. Such finding would reveal an unsuspected malleability of the A/B compartments linked with the symbiotic status, and hamper support for the existence of static two-speed genome in plant associated fungi (84, 85).

Conclusions

By combining long-reads with Hi-C sequencing we demonstrate that model mycorrhizal symbionts all carry 33 chromosomes. Their gene content and size of varies significantly, even between strains still referred by others as "clones" (86), supporting the notion that AMF have pangenomes. 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. Although this pattern is reminiscent of the "two-speed" genome structure of some fungal plant pathogens, we also find primary evidence that the A/B genome structure may fluctuate with symbiotic conditions.

To conclude, this work represents a springboard that researchers can now use to generate "gene expression atlases" of this model plant symbiont under different environmental conditions, and acquire chromosome-level data from other fungal species. This study also opens avenues to identify the molecular mechanisms that generate and modify chromosome compartments during mycorrhizal symbioses and other host-microbe interactions. It will be now interesting to examine how the present results extrapolate to AMF dikaryons, in particular how these strains compartmentalize their co-existing parental genomes, and how/if these also significantly differ in gene content and size. Phasing parental genomes with Hi-C data will be needed to fully address these questions, and to determine how co-existing genomes regulate nuclear dynamics in these strains (11, 87).

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. Our research is funded by the Discovery program of the Natural Sciences and Engineering Research Council (RGPIN2020-05643), a 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 S1: *R. irregularis* C2, A1, B3 and 4401 strain chromosomes. The length of all 33 chromosomes is indicated in the karyoplots. Telomere repeats, MAT locus and small subunit

rRNA locations are marked. Gene densities are shown in red, and repeat densities are shown in blue color.

Figure S2: Bubble plot showing 10 most abundant transposable elements and their distribution in five assemblies of *R. irregularis* strains.

Figure S3: Gene orthologies and Pfam domain numbers in five strains of *R. irregularis*. a) Protein orthologs of A1, B3, C2, DAOM-197198 and 4401 strains. **b) Pfam protein domain abundance comparison between five *R. irregularis* strains.** After the 100 most abundant Pfam protein domains were selected, the frequency values were transformed using a t-test. This way, relative enrichment (red) and relative depletion (blue) were shown.

Figure S4: Hierarchical clustering of *k*-mer distance estimations between chromosomes of the strains.

Figure S5: Heat map showing Hi-C contact frequencies between 33 chromosomes for strain DAOM-197198, C2, A1, B3 and 4401. Regions having higher contact frequencies are found in close contact in the nucleus, whereas regions having fewer contact frequencies are far away from each other.

Figure S6: Heat map showing Hi-C contact probabilities of individual chromosomes of strain DAOM-197198. Chromosomal regions having higher contact probabilities are found in close contact with each other, whereas regions having fewer contact probabilities are far away from each other. PCA eigenvectors were calculated for 10kb intervals and were plotted under the chromosome contact maps.

Figure S7: Heat map showing Hi-C contact probabilities of individual chromosomes of strain C2. Chromosomal regions having higher contact probabilities are found in close contact with each other, whereas regions having fewer contact probabilities are far away from each other. PCA eigenvectors were calculated for 10kb intervals and were plotted under the chromosome contact maps.

Figure S8: Heat map showing Hi-C contact probabilities of individual chromosomes of strain A1 Chromosomal regions having higher contact probabilities are found in close contact with each other, whereas regions having fewer contact probabilities are far away from each

other. PCA eigenvectors were calculated for 10kb intervals and were plotted under the chromosome contact maps.

Figure S9: Heat map showing Hi-C contact probabilities of individual chromosomes of strain B3 Chromosomal regions having higher contact probabilities are found in close contact with each other, whereas regions having fewer contact probabilities are far away from each other. PCA eigenvectors were calculated for 10kb intervals and were plotted under the chromosome contact maps.

Figure S10: Heat map showing Hi-C contact probabilities of individual chromosomes of strain 4401 Chromosomal regions having higher contact probabilities are found in close contact with each other, whereas regions having fewer contact probabilities are far away from each other. PCA eigenvectors were calculated for 10kb intervals and were plotted under the chromosome contact maps.

Figure S11: Strain-specific chromosomal differentiation events can be observed in Hi-C contact maps. Differentiation events that can be observed on contact maps and synteny plots were shown with an orange arrow. Repeat densities of the chromosomes are shown in blue color, and the synteny blocks are shown in purple. **a) Strain-specific repeat expansion found in chromosome 8 of strain 4401.** A) A repeat expansion event can be observed in chromosome 8 of 4401 strain, between regions 1Mb and 3Mb. As the same repeats get expanded, the contact of Hi-C reads in that region increase, leading to the increased brightness. B) Synteny blocks between chromosome 8 from DAOM-197198 and 4401 shows absence of synteny blocks in the repeat expansion region. In the regions showing no synteny between the two strains, DAOM-197198 strain has 157 genes and 1240 repeats whereas 4401 strain has 389 genes and 2799 repeats. **b) Strain specific inversion event in chromosome 18 of strain C2.** A) An inversion event can be observed in chromosome 18 of C2 strain, around 2Mb region. This differentiation event also causes an abnormality in the contact map around that region. B) Synteny blocks between chromosome 18 from DAOM-197198 and C2 show multiple inversion events. **c) Strain specific inversion event in chromosome 17 of strain A1.** A) An inversion event can be observed in chromosome 17 of A1 strain, before 1 Mb region. This differentiation event also causes an abnormality in the contact map around that region. B) Synteny blocks between chromosome 17 from DAOM-197198 and A1 show multiple inversion events. **d) Strain specific**

inversion event in chromosome 18 of strain B3. A) An inversion event can be observed in chromosome 18 of B3 strain, around 2Mb region. This differentiation event also causes an abnormality in the contact map around that region. B) Synteny blocks between chromosome 18 from DAOM-197198 and B3 show multiple inversion events.

Figure S12: Distribution of core and dispensable genes in compartments A and B. Protein orthologs of A1, C2 and DAOM-197198 strains were identified with FastOrtho (57), using following parameters: 50% identity and 50% coverage. The genes were then grouped into A and B compartments. Core and core duplicated gene numbers were added up in each compartment to represent the "core group" count, and dispensable, specific and specific duplicated gene numbers were added up to make up the "dispensable group" count.

Figure S13: A) For each A/B compartment, the median methylation frequencies are collected and their distribution is plotted**.** The A compartment display higher median methylation frequencies than the B compartment. Highly methylated CpG sites (median methylation frequency > 80%) occur more frequently in the A compartment. **B)** CpG methylation frequencies of transposable elements in compartments A and B. Individual CpG methylation frequencies of transposable elements found in compartments A and B are shown as green dots. Methylation medians are shown as red dots.

Figure S14: Example of inversions in synteny blocks between chromosome 13 of DAOM-197198, C2 and A1 strains. The colors on the karyoplots represent the compartments. Compartment A is represented by blue color, whereas compartment B is represented by red color. Synteny blocks are created by MCScanx(88), and inversions are highlighted. Orange links show inverted synteny blocks between DAOM-197198 and A1, and green links show inverted synteny blocks between DAOM-197198 and C2. Majority of the inversions are observed in compartment B.

Table S1: Location and number of ribosomal RNA operons within and across strains analyzed in this study.

Table S2: A/B compartment properties in the model strain DAOM-197198. The A-compartment is more gene-dense, while the B-compartment is more repeat-rich. However, the B-compartment has significantly more secreted proteins, candidate apoplastic effectors and

candidate cytoplasmic effectors than the A-compartment (Fisher's exact test: * indicates $p < 0.05$).

Table S3: Comparison of A and B compartment similarity in DAOM-197198, C2 and A1 strains. Overall, A compartment is more conserved with higher similarity between strains compared to B compartment.

References

1. J. W. Spatafora, *et al.*, A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**, 1028–1046 (2016).

2. F. Lutzoni, *et al.*, Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat. Commun.* **9**, 1–11 (2018).

3. N. Corradi, P. Bonfante, The Arbuscular Mycorrhizal Symbiosis: Origin and Evolution of a Beneficial Plant Infection. *PLoS Pathog.* **8**, e1002600 (2012).

4. F. M. Martin, S. Uroz, D. G. Barker, Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science (80-.).* **356** (2017).

5. P. Bonfante, I. A. Anca, Plants, mycorrhizal fungi, and bacteria: A network of interactions. *Annu. Rev. Microbiol.* **63**, 363–383 (2009).

6. P. M. Delaux, S. Schornack, Plant evolution driven by interactions with symbiotic and pathogenic microbes. *Science (80-.).* **371** (2021).

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

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

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

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

11. V. Kokkoris, *et al.*, Host identity influences nuclear dynamics in arbuscular mycorrhizal

fungi. *Curr. Biol.* **31**, 1531-1538.e6 (2021).

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

13. K. Lin, *et al.*, Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an Endomycorrhizal Fungus. *PLoS Genet.* **10** (2014).

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

15. E. Morin, *et al.*, Comparative genomics of Rhizophagus irregularis, R. cerebriforme, R. diaphanus and Gigaspora rosea highlights specific genetic features in Glomeromycotina. *New Phytol.* **222**, 1584–1598 (2019).

16. M. Malar C, *et al.*, The genome of Geosiphon pyriformis reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Curr. Biol.* **31**, 1570-1577.e4 (2021).

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

18. C. G. P. McCarthy, D. A. Fitzpatrick, Pan-genome analyses of model fungal species. *Microb. Genomics* **5** (2019).

19. X. Li, *et al.*, Complete genome sequence of the sesame pathogen Ralstonia solanacearum strain SEPPX 05. *Genes and Genomics* **40**, 657–668 (2018).

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

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

22. J. F. Denton, *et al.*, Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLoS Comput. Biol.* **10** (2014).

23. M. Montoliu-Nerin, *et al.*, Building de novo reference genome assemblies of complex eukaryotic microorganisms from single nuclei. *Sci. Rep.* **10** (2020).

24. Q. Szabo, F. Bantignies, G. Cavalli, Principles of genome folding into topologically associating domains. *Sci. Adv.* **5** (2019).

25. J. Dekker, E. Heard, Structural and functional diversity of Topologically Associating Domains. *FEBS Lett.* **589**, 2877–2884 (2015).

26. A. Pombo, N. Dillon, Three-dimensional genome architecture: Players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).

27. R. Kempfer, A. Pombo, Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).

28. D. J. Winter, *et al.*, Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus Epichloë festucae. *PLoS Genet.* **14** (2018).

29. J. M. Galazka, *et al.*, Neurospora chromosomes are organized by blocks of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. *Genome Res.* **26**, 1069–1080 (2016).

30. A. D. Schmitt, M. Hu, B. Ren, Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.* **17**, 743–755 (2016).

31. H. Stockinger, C. Walker, A. Schüßler, "Glomus intraradices DAOM197198", a model fungus in arbuscular mycorrhiza research, is not Glomus intraradices. *New Phytol.* **183**, 1176–1187 (2009).

32. A. M. Koch, *et al.*, High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2369–2374 (2004).

33. N. Corradi, M. Hijri, L. Fumagalli, I. R. Sanders, Arbuscular mycorrhizal fungi (Glomeromycota) harbour ancient fungal tubulin genes that resemble those of the chytrids (Chytridiomycota). *Fungal Genet. Biol.* **41**, 1037–1045 (2004).

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

35. S. Koren, *et al.*, Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

36. R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly

from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).

37. B. J. Walker, *et al.*, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, 112963 (2014).

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

39. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

40. M. J. Roach, S. A. Schmidt, A. R. Borneman, Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 1–10 (2018).

41. Arima Genomics, 01_mapping_arima.sh (2019) (June 8, 2021).

42. J. Ghurye, *et al.*, Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput. Biol.* **15**, e1007273 (2019).

43. J. Ghurye, M. Pop, S. Koren, D. Bickhart, C. S. Chin, Scaffolding of long read assemblies using long range contact information. *BMC Genomics* **18**, 527 (2017).

44. T. Maeda, *et al.*, Evidence of non-tandemly repeated rDNAs and their intragenomic heterogeneity in Rhizophagus irregularis. *Commun. Biol.* **1** (2018).

45. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

46. S. El-Gebali, *et al.*, The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

47. T. N. Petersen, S. Brunak, G. Von Heijne, H. Nielsen, SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).

48. A. Krogh, B. Larsson, G. Von Heijne, E. L. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

49. J. Sperschneider, P. N. Dodds, EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *bioRxiv*, 454080 (2021).

50. J. M. Flynn, *et al.*, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 9451–9457 (2020).

51. A. F. A. Smit, R. Hubley, P. Green, No Title. *RepeatMasker Open-4.0.*

52. J. M. Hancock, M. J. Bishop, "HMMer" in *Dictionary of Bioinformatics and Computational Biology*, (John Wiley & Sons, Ltd, 2004) https:/doi.org/10.1002/9780471650126.dob0323.pub2 (February 5, 2021).

53. W. Li, *et al.*, The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43**, W580–W584 (2015).

54. B. D. Ondov, *et al.*, Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 1–14 (2016).

55. G. Marçais, *et al.*, MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14** (2018).

56. S. Kurtz, *et al.*, Versatile and open software for comparing large genomes. *Genome Biol.* **5**, 12 (2004).

57. A. R. Wattam, *et al.*, PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42**, D581 (2014).

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

59. F. Ramírez, *et al.*, High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 1–15 (2018).

60. N. Servant, *et al.*, HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16** (2015).

61. J. Wolff, *et al.*, Galaxy HiCExplorer: A web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46**, W11–W16 (2018).

62. J. Wolff, *et al.*, 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 Res.* **48**, W177–W184 (2020).

63. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

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

65. N. Corradi, *et al.*, Gene copy number polymorphisms in an arbuscular mycorrhizal fungal population. *Appl. Environ. Microbiol.* **73**, 366–369 (2007).

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

67. B. P. H. J. Thomma, *et al.*, Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genet. Biol.* **90**, 24–30 (2016).

68. M. Spielmann, D. G. Lupiáñez, S. Mundlos, Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).

69. J. Sperschneider, *et al.*, The stem rust fungus Puccinia graminis f. sp. tritici induces waves of small RNAs with opposing profiles during wheat infection. *bioRxiv*, 469338 (2018).

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

71. L. Kamel, *et al.*, The comparison of expressed candidate secreted proteins from two arbuscular mycorrhizal fungi unravels common and specific molecular tools to invade different host plants. *Front. Plant Sci.* **8** (2017).

72. H. Zheng, W. Xie, The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).

73. I. Jerković, Q. Szabo, F. Bantignies, G. Cavalli, Higher-Order Chromosomal Structures Mediate Genome Function. *J. Mol. Biol.* **432**, 676–681 (2020).

74. A. J. Lea, *et al.*, Genome-wide quantification of the effects of DNA methylation on human gene regulation. *Elife* **7** (2018).

75. A. A. Dallaire, Transcriptional activity and epigenetic regulation of transposable elements in the symbiotic fungus Rhizophagus irregularis. *bioRxiv*, 2021.03.30.436303 (2021).

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

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

78. T. D. Bruns, N. Corradi, D. Redecker, J. W. Taylor, M. Öpik, Glomeromycotina: what is a species and why should we care? *New Phytol.* **220**, 963–967 (2018).

79. H. S. Garmaroodi, M. Taga, Duplication of a conditionally dispensable chromosome carrying pea pathogenicity (PEP) gene clusters in Nectria haematococca. *Mol. Plant-Microbe Interact.* **20**, 1495–1504 (2007).

80. I. Vlaardingerbroek, B. Beerens, S. M. Schmidt, B. J. C. Cornelissen, M. Rep, Dispensable chromosomes in Fusarium oxysporum f. Sp. lycopersici. *Mol. Plant Pathol.* **17**, 1455–1466 (2016).

81. S. Bertazzoni, *et al.*, Accessories Make the Outfit: Accessory Chromosomes and Other Dispensable DNA Regions in Plant-Pathogenic Fungi. *Mol. Plant-Microbe Interact.* **31**, 779–788 (2018).

82. J. M. Plett, F. Martin, Poplar root exudates contain compounds that induce the expression of MiSSP7 in Laccaria bicolor. *Plant Signal. Behav.* **7** (2012).

83. J. L. Soyer, *et al.*, Epigenetic Control of Effector Gene Expression in the Plant Pathogenic Fungus Leptosphaeria maculans. *PLoS Genet.* **10**, e1004227 (2014).

84. L. Frantzeskakis, S. Kusch, R. Panstruga, The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol. Plant Pathol.* **20**, 3–7 (2019).

85. D. E. Torres, U. Oggenfuss, D. Croll, M. F. Seidl, Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biol. Rev.* **34**, 136–143

(2020).

86. T. Wyss, F. G. Masclaux, P. Rosikiewicz, M. Pagni, I. R. Sanders, Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus Rhizophagus irregularis. *ISME J.* **10**, 2514–2526 (2016).

87. E. U. Serghi, *et al.*, Homo- and Dikaryons of the Arbuscular Mycorrhizal Fungus Rhizophagus irregularis Differ in Life History Strategy. *Front. Plant Sci.* **12**, 1544 (2021).

88. Y. Wang, *et al.*, MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).