Stochastic nuclear organization and host-dependent allele contribution in Rhizophagus irregularis

Erik Limpens (erik.limpens@wur.nl)
Wageningen University and Research Wageningen Plant Research https://orcid.org/0000-0002-9668-4085

Jelle van Creij
Wageningen University and Research Wageningen Plant Research

Ben Auxier
Wageningen University and Research Wageningen Plant Research

Jianyong An
Wageningen University and Research Wageningen Plant Research

Raúl Y. Wijfjes
Wageningen University and Research Wageningen Plant Research

Claudia Bergin
Uppsala University: Uppsala Universitet

Anna Rosling
Uppsala University: Uppsala Universitet

Ton Bisseling
Wageningen University and Research Wageningen Plant Research

Zhiyong Pan
Huazhong Agriculture University: Huazhong Agricultural University

Research Article

Keywords: arbuscular mycorrhiza, heterokaryote, recombination, parasexual, single nucleus sequencing, symbiosis, Rhizophagus irregularis, PacBio SMRT sequencing

Posted Date: February 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1113831/v1

Stochastic nuclear organization and host-dependent allele contribution in *Rhizophagus irregularis*

Jelle van Creij[1], Ben Auxier[2], Jianyong An[1,3], Raúl Y. Wijfjes[4,5], Claudia Bergin[6], Anna Rosling[7], Ton Bisseling[1,3], Zhiyong Pan[8], Erik Limpens[1]

1. Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen University & Research, Droevendaalsesteeg 1, Wageningen, The Netherlands
2. Laboratory of Genetics, Department of Plant Sciences, Wageningen University & Research, Droevendaalsesteeg 1, Wageningen, The Netherlands
3. Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing University of Agriculture, Beijing, 102206, China.
4. Laboratory of Bioinformatics, Department of Plant Sciences, Wageningen University & Research, Droevendaalsesteeg 1, Wageningen, The Netherlands
5. Current affiliation: Faculty of Biology, Ludwig Maximilian University of Munich, München, Germany
6. Department of Cell and Molecular Biology, Uppsala University, and Microbial Single Cell Genomics Facility, Science for Life Laboratory, Uppsala, Sweden.
7. Department of Ecology and Genetics, Uppsala University, Norbyvägen 18D, SE-75236, Uppsala, Sweden.
8. Key Laboratory of Horticultural Plant Biology (Ministry of Education), Key Laboratory of Horticultural Crop Biology and Genetic Improvement (Central Region, Ministry of Agriculture), College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, P.R. China.

Keywords: arbuscular mycorrhiza, heterokaryote, recombination, parasexual, single nucleus sequencing, symbiosis, *Rhizophagus irregularis*, PacBio SMRT sequencing

Abstract

Background

Arbuscular mycorrhizal (AM) fungi are arguably the most important symbionts of plants, offering a range of benefits to their hosts. However, the provisioning of these benefits does not appear to be uniform among AM fungal individuals, with genetic variation between fungal symbionts having a substantial impact on plant performance. Interestingly, genetic variation has also been reported within fungal individuals, which contain millions of haploid nuclei sharing a common cytoplasm. In the model AM fungus, *Rhizophagus irregularis*, several isolates have been reported to be dikaryotes, containing two genetically distinct types of nuclei recognized based on their mating-type (MAT) locus identity. However, their extremely coenocytic nature and lack of a known single nucleus stage has raised questions on the origin, distribution and dynamics of this genetic variation.

Results

Here we performed DNA and RNA sequencing at the mycelial individual, single spore and single nucleus levels to gain insight into the dynamic genetic make-up of the *R. irregularis* C3 isolate. This isolate is thought to be clonally related to the recently sequenced dikaryon-like A4 isolate, which both were isolated ~ 20 years ago from the same field in Switzerland. Our analyses reveal that both isolates vary considerably in their nuclear behavior. Parallel spore and root culture batches showed widely variable ratios of two main nucleotypes in C3. Additionally, numerous polymorphisms were found that deviated significantly from the distribution of the two main nucleotypes in C3. No consistent host effects on nucleotype ratio after multiple rounds of subculturing were observed. Instead, we found a major effect of host plant-identity on allele-specific expression in C3.

Conclusion

Our analyses indicate a much more dynamic/variable genetic organization in *R. irregularis* than previously assumed. Seemingly random fluctuations in nucleotype ratio's upon spore formation,

recombination events, high variability of non-tandemly repeated rDNAs and host-dependent allele-specific expression all add levels of variation that may contribute to the evolutionary success of these widespread symbionts.

Introduction

Fungi belonging to the Glomeromycotina subphylum of the Mucoromycota are globally distributed soil fungi that form an endosymbiosis with the vast majority of land plants [1]. These so-called arbuscular mycorrhizal (AM) fungi rely on their interaction with plants to complete their life cycle. During colonization of plant roots, they form highly branched structures called arbuscules inside root inner cortex cells, where mineral nutrients such as phosphate and nitrogen are exchanged for sugars and fatty acids from the plant [2]. This symbiosis originated more than 400 million years ago and has since been maintained in the vast majority of plants, highlighting its importance in natural ecosystems [3]. Currently, around 315 AM fungal species have been described, however the species concept for these enigmatic fungi is not well defined [4]. Significant intraspecific genetic variation has been observed but evidence for sexual reproduction remains elusive. How the genetic organization of these important fungi contributes to the evolutionary success of this key symbiosis is an important and highly debated question [5,6]. Large variations in the symbiotic performance, often referred to as mycorrhizal growth response, of different isolates or even between strains derived from single spores from a fungal individual has been reported [7,8]. What determines this variation in mycorrhizal growth response, i.e. how much growth benefit a plant has from interacting with a certain fungus, remains unknown. An important first step to understanding the mycorrhizal response is understanding if the genetic organization of AM fungi adapts to different environments and plant hosts, impacting their growth.

Among fungi, AM fungi have relatively large genome sizes (~150 – 750 Mb) and are rich in transposable elements [6]. They form multinucleate spores and their hyphal network contains millions of haploid nuclei that occupy a shared cytoplasm [5,6,9]. Such coenocytic hyphae generally lack cross-walls and nuclei can flow freely from the hyphae into the spores as they form and grow [10,11]. As a result, spores contain hundreds of nuclei and there is no known single nucleus stage that generates the next generation. Although other fungi with multinucleate hyphae and spores are known [12,13], to our

knowledge the extremely large coenocytic nuclei number and apparent lack of a single nucleus stage is unique to AM fungi.

Increasing the uncertainty about the genetics of AM fungi, sexual structures have never been observed [6]. Therefore, historically AM fungi were thought to propagate asexually, raising questions about their ability to purge deleterious mutations and to generate genetic variation required for adaptation. One mechanism proposed a large variety of genetically diverse nuclei in fungal individuals, and subsequent selection on individual nuclei [5,14]. However, the availability of various whole genome sequences from different AM fungi has somewhat challenged this view, revealing much lower intra-organismal genetic variation than previously assumed [15-21]. Furthermore, AM fungi were found to contain a full complement of the core genes required for meiosis [15,22,23]. A putative mating-type (MAT) locus, consisting of two HD-like genes, has been identified in *Rhizophagus irregularis,* consistent with a bipolar mating system [18]. Whole genome sequencing together with single nucleus sequencing of various *R. irregularis* isolates revealed that some were in fact monokaryotic (ie. containing genetically very similar nuclei representing one nucleotype with a single MAT allele), while others (such as isolates A4 and A5) appeared to be dikaryotic (ie. two different nucleotypes carrying two distinct MAT alleles). Furthermore, allele frequency analyses indicated a mostly 1:1 ratio of the two nucleotypes in the two dikaryotic strains studied [18]. A recent study by the same group suggested that in dikaryotic strains the ratio of the two nucleotypes may shift in response to host plant identity [24].

Hyphal fusion and exchange of nuclei has been observed between closely related AM fungi [25,26]. However, the genetic factors that control the vegetative compatibility between AM fungi are not known and the fate of nuclei that are exchanged is also not yet well understood. Potential inter-nucleus recombination between nuclei with distinct MAT loci was recently suggested based on single nucleus sequencing in dikaryotic strains [19,30], although the extent by which this occurs is still debated [31]. Inter-nucleus recombination could suggest that nuclear fusion followed by meiosis does occur, although alternative explanations such as a parasexual cycle where nuclei fuse and undergo mitotic

recombination followed by a return to the haploid state through the (stochastic) loss of chromosomes cannot be excluded [32]. To date, no diploid or aneuploid nuclei have been observed in AM fungi.

To better understand the organization of intra-strain genetic variation in AM fungi, we studied the genetic organization of the presumed dikaryote-like *R. irregularis* C3 isolate [33,34]. By using a combination of culture meta-genome, single spore and single nucleus sequencing we generated a high quality genome assembly for the C3 isolate. Allele frequency analyses of the mating locus revealed an unequal ratio of two main nucleotypes. Interestingly, allele frequency distribution varied significantly between single spores and between different spore and hyphal batches, suggesting a stochastic segregation of nuclei. Contrary to results from other *R. irregularis* isolates, such as its presumed clonal sister strain A4, no consistent host effects on allele frequency distributions were found after multiple rounds of subculturing. Instead, transcriptome data showed a strong host-dependent effect on allele-specific expression in the interaction of C3 with different host plants.

Results

Characterising intragenomic variation

Genome assembly

The *R. irregularis* C3 isolate was initially chosen because of its reported relative high level of genetic variation, based on RADseq data [34], which in hindsight was overestimated due to lack of a good reference genome [35]. To characterize the genetic variation in this isolate we first generated a C3 reference genome, using a combination of PacBio and Illumina sequencing on genomic DNA extracted from a large number of spores and hyphae from axenic *Daucus carota* root cultures (Fig. 1; Table 1). This resulted in an assembly (RirC3; Genbank BioProject ID PRJNA747641) comprising 1380 contigs and representing a total length of 155 Mbp (Table 1, Fig. 1); a genome size similar to previous estimates for the genome length in *R. irregularis* strains [6]. A representation of the 10 longest contigs covering 9.9 Mb, depicting the distribution of repeats, coding regions and SNP density, is shown in Figure 1.

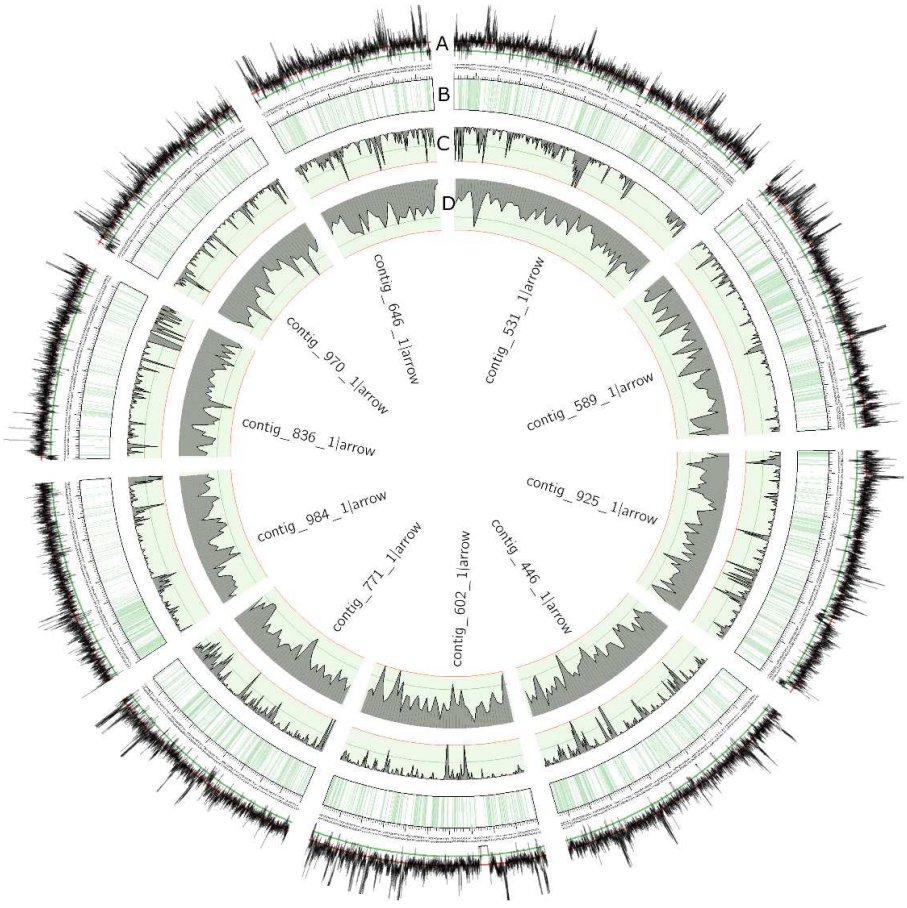


Figure 1. Circos diagram of the ten largest contigs of the RirC3 assembly, representing 9.9 Mb. **A**: Mapping depth of C3 Illumina reads, green line = 50, red line = 100. **B**: Physical map of the contigs, with coding regions coloured green. **C**: SNP density, green line = 10, red line = 20. **D**: Repeat density, green line = 75, red line = 100.

Table 1: RirC3 genome assembly overview.

Total assembly length	155,051,422
Contigs	1380
Gaps	2
GC content	27.94%
Longest contig length	1,364,683
Average contig length	112.356
Contig N50 length	263.815
Contig L50	174
Contig N90 length	63.299
Contig L90	605
Total repeat length	71,617,038
Percentage of genome	46.19%
Total number of SNPs	121.109
Non-synonymous SNPs	10.677
Stop codon gains	381
Frameshift gains	487
SNP frequency	0.78/kb
SNP density outside repeats	0.59/kb
Number of predicted genes	27,181
PFAM domains	3,714
BUSCO (fungi_odb10)	84.3% (639/758)
Complete and single-copy	82.6% (626/758)
Complete and duplicated	1.7% (13/758)
Fragmented	1.2% (9/758)
Missing	14.5% (110/758)

Repeated regions, including transposable elements, represented 46% of the genome (71 Mb). These repeat regions appear to be randomly distributed over the genome and the majority remain unclassified (Fig. 1). The genome assembly contained 85% of the BUSCO (fungi_odb10) gene set, which is similar to the completeness observed for the high quality *R. irregularis* DAOM197198 (Rir17; [17] genome. BUSCO genes that were not found include FATTY ACID SYNTHASE I and other genes reported to be consistently lost in the *R. irregularis* genome (Additional File 1), correlating with their obligate biotrophic lifestyle [36]. Nearly all BUSCO genes were found in a single copy, indicating a low level of haplotig duplication from the two nucleotypes. The genome (RirC3) assembly was further annotated using the FunAnnotate pipeline, specialized for fungal genome annotation [37], resulting in 27181 predicted gene models.

45S rDNA organization

R. irregularis DAOM197198 (Rir17) was reported to contain an atypical non-tandemly repeated organization of the 45S rDNA locus, consisting of 10 or 11 copies [17]. Similarly, RirC3 contains only eight 45S rDNA copies that lack a tandem organization. Four of these copies were located on separate contigs; the other four were found in two pairs, separated over 50 kb apart on separate contigs.

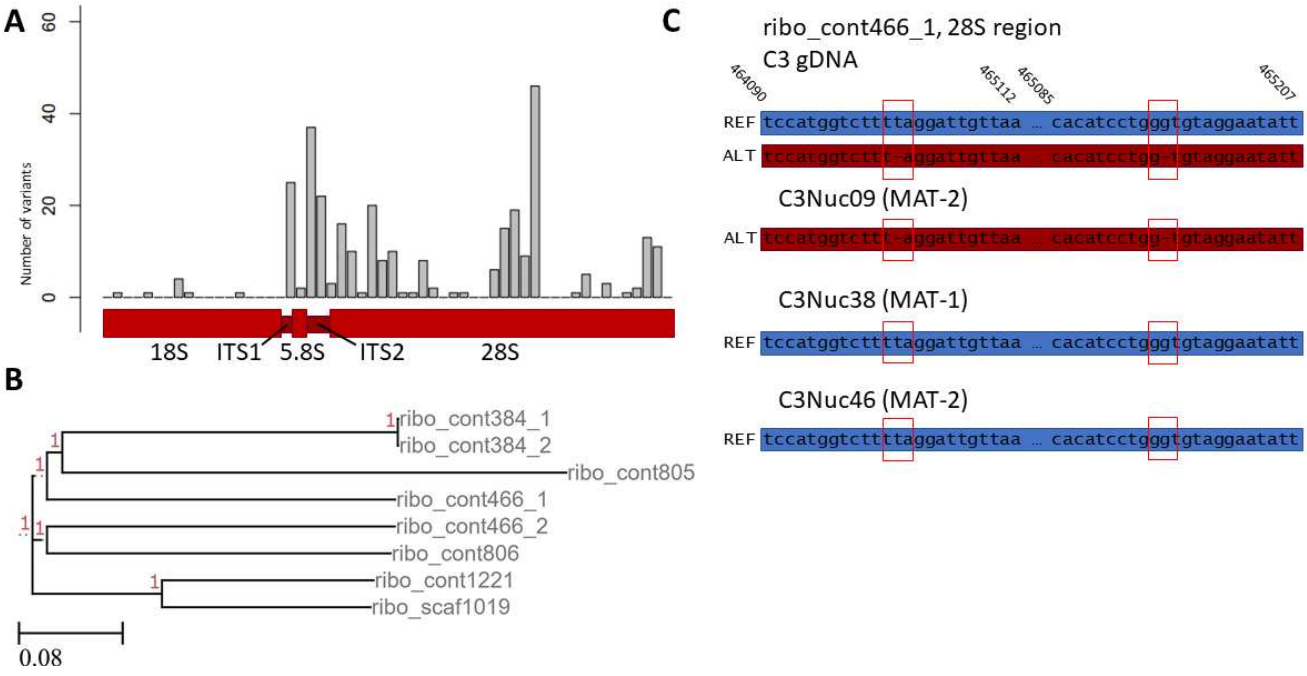


Figure 2. Polymorphisms found in the RirC3 45S rDNA locus. **A**: Graph showing the amount of genetic variation between 45S rDNA copies. **B**: Phylogenetic representation based on of multiple sequence alignment (1000 bootstraps) of the eight 45S rDNA copies. No copies were identical. The names of the samples correspond to which contig they were found on (e.g. ribo_cont466_1 was the first copy on contig_466 of the RirC3 assembly). Red numbers indicate support values. **C**: Example of additional polymorphisms in the rDNA sequence distributed over different nuclei. The reference and alternate alleles for a 28S subregion of ribo_contig466_1 in the C3 gDNA and 3 individually sequenced nuclei (nuc9, nuc38 and nuc46) are shown.

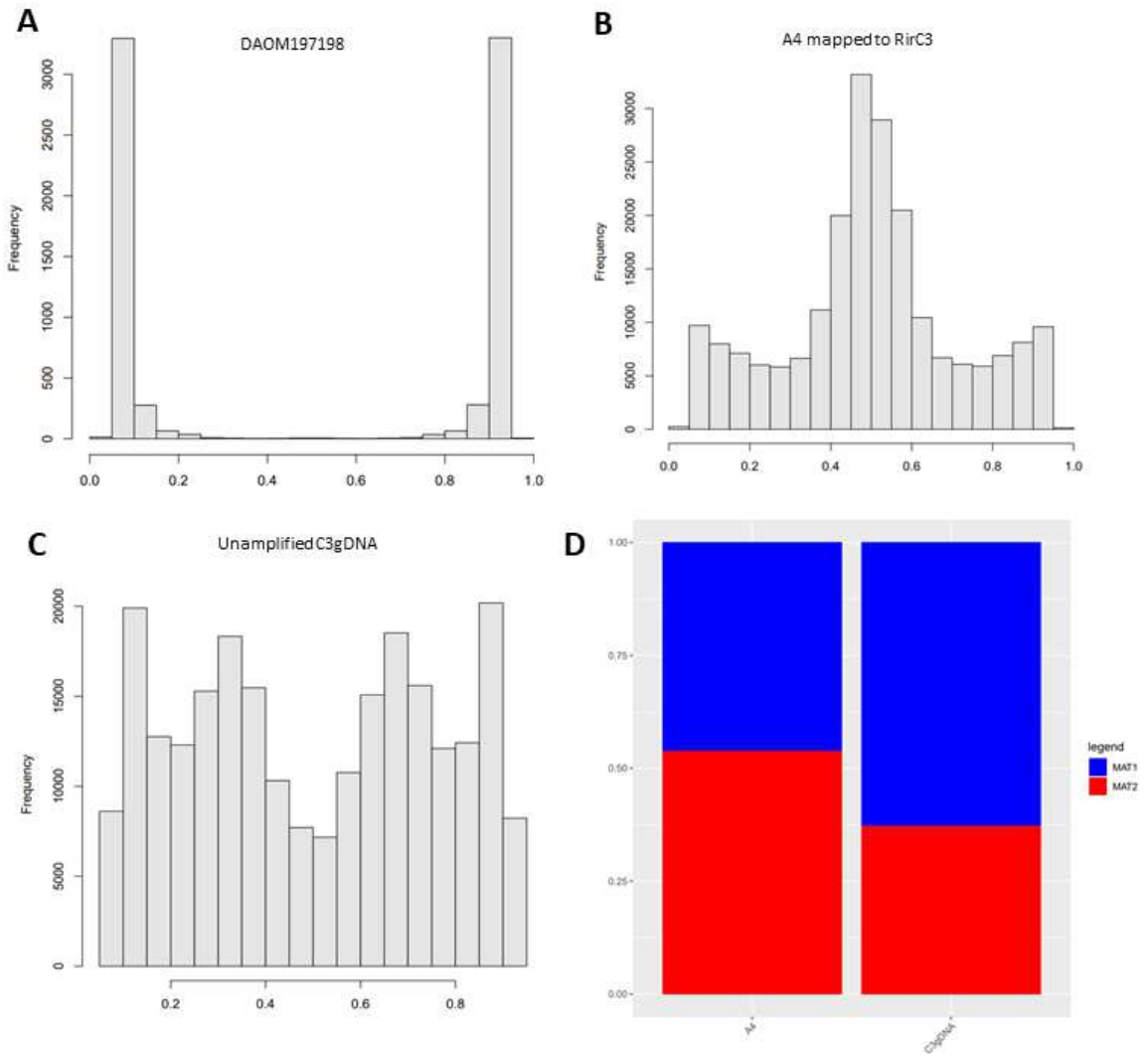
Alignment of the sequences of these copies showed significant variation between the different loci, each consisting of 18S rDNA, intergenic spacer region 1 (ITS1), 5.8S rDNA, ITS2, and 28S rDNAs (Fig. 2a,b). When assessing the sequencing depth at these 45S rDNA sequences, we found no increased coverage that would suggest a collapse of assembled sequences as would be expected in the case of several highly conserved copies. Upon analyzing the number of polymorphisms between the 45S rDNA copies, we identified 31 SNP's within four of the 45S rDNA contigs (Additional File 2). These data support the relative high heterogeneity of *R. irregularis* 45S rDNA copies, which has been suggested to potentially modulate the translational activity of different ribosomes [17]. Single nucleus sequencing (see below) showed that different nuclei indeed encode distinct rDNA alleles, confirming the observed heterogeneity in the assembly (Fig. 2c).

Allelic variation

To investigate the genome-wide level of genetic variation, SNP calling was performed using Freebayes based on Illumina sequencing reads from DNA isolated from a large collection of root culture plates, referred to as meta-genome (C3gDNA). SNPs were filtered based on a coverage within the 25th percentile from the average mapping depth, with at least 10 observations of the alternate allele. With these settings 121.109 SNPs were found (Additional File 3), giving a SNP density of 0.79 SNPs/kb. After

175 removing SNPs that were located inside repetitive regions 0.59 SNPs/kb remained. 10.677 SNPs
176 represented non-synonymous SNPs in the predicted protein coding genes. Allele frequency
177 distribution analyses confirmed the homokaryotic nature of the DAOM197198 isolate and the reported
178 50:50 distribution of allelic variations in the A4 isolate (Fig. 3a,b). However, the allele frequency
179 distribution in the C3 meta-genome sample did not match the predicted 50:50 ratio that was visible
180 for A4. Instead, two peaks were found corresponding to 33% and 67% allele frequencies for C3 (Fig.
181 3c). Such an allele frequency distribution is typically found in triploid genomes [38]. The observed 2:1
182 SNP ratio had a consistent genome-wide distribution, ruling out that this distribution was caused by
183 local aneuploidy.

184



185

Figure 3. Allele frequencies in different *Rhizophagus irregularis* isolates. **A**: DAOM197198, mapped to Rir17 [17]. **B**: A4 reads [from 18] mapped to the RirC3 assembly. **C**: C3 Illumina reads mapped to the RirC3 assembly. **D**: MAT locus proportions based on coverage of the MAT loci in A4 and C3 Illumina reads mapped against the RirC3 assembly.

To determine whether such unequal allele frequencies were consistent between C3 cultures, gDNA of another batch of C3 root culture plates was also sequenced. Whole genome amplification (WGA) was used to generate sufficient DNA before sequencing of this sample (C3_WGA1, Fig 4a). To rule out any artefacts introduced by WGA, the original C3gDNA sample used for the assembly was also amplified (C3_WGA2) and sequenced (Fig 4b). To further monitor the reproducibility of the whole genome amplification procedure with respect to SNP frequencies, multiple WGA replicates were included for both meta gDNA samples (meta refers to the use of a large number of spores and mycelium) (Fig 4c). Principal component analysis (PCA) of allele frequencies showed that the whole genome amplification did not introduce much variation in allele frequencies between technical replicate samples as seen by the tight clustering of these samples (Fig. 4d). This indicated that whole genome amplification did not cause a bias in the allele frequencies of the respective samples. However, it also showed that the two meta DNA samples, isolated from different batches of root culture plates, differed in allele frequencies. The C3_WGA2 samples showed allele frequency peaks at ~33% and 67% in line with the allele frequency distribution in the unamplified C3gDNA, while the others (C3_WGA1 replicates) showed a rather broad peak around 50% suggestive of a 1:1 nucleotype ratio (Fig. 4a,b, Additional File 9: Fig. S1).

These analyses suggested that different batches can differ in their nucleotype ratio's. Two additional DNA samples, each from 50 spores collected from two other root culture plates (labelled C3_spores_1 and C3_spores_2), were sequenced after whole genome amplification. These again showed divergent allele frequency distributions based on both genome-wide allele frequencies and MAT allele ratio's (Fig. 4c,e).

To further investigate the nucleotype ratio's we searched for the presumed MAT loci [18]. We identified two MAT loci identical to the MAT-1 and MAT-2 sequences reported for A4 [18]. Read mapping to these loci showed similar ratios consistent with the genome wide SNP analyses; approximately 1:1 in the C3-WGA2 reads and 2:1 in the C3_WGA1 reads (Fig 4e). Variable nucleotype frequencies were also observed based on MAT allele ratios determined by ddPCR on multiple unamplified DNA samples collected from different root culture plates (Fig. 4f), confirming that the observed variation was not caused by the whole genome amplification.

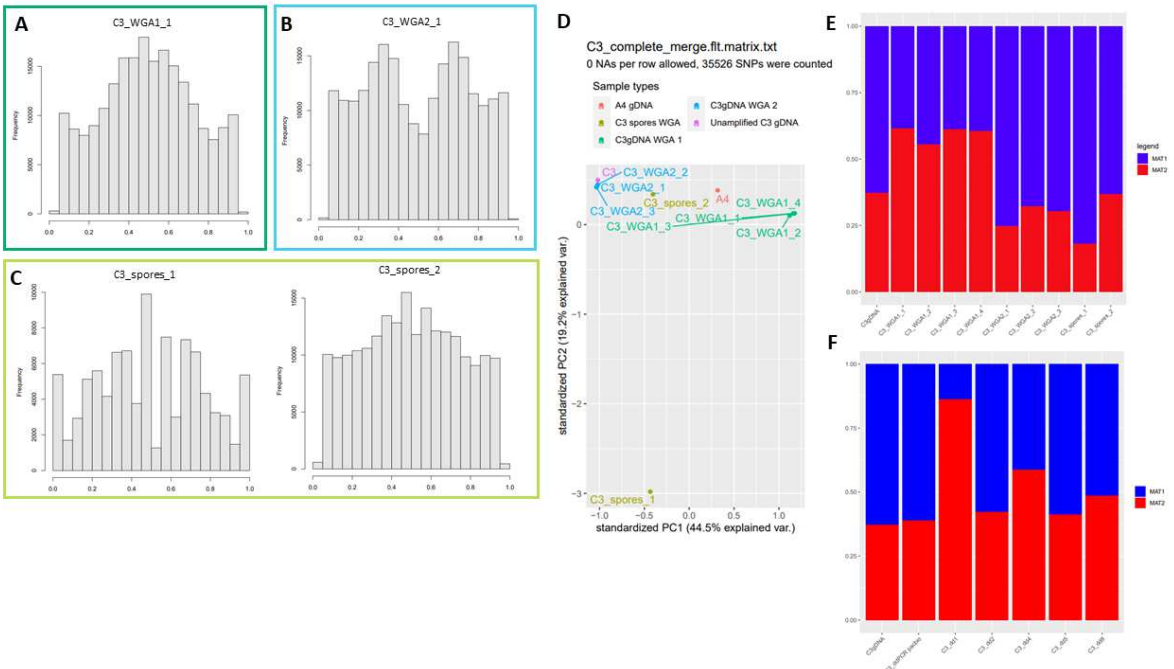


Figure 4. Allele frequencies of different C3 colonies. **A**: Allele frequency distribution of C3 genomic DNA that was amplified from the sample used for Illumina sequencing. **B**: Allele frequency distribution of amplified C3 genomic DNA from a previously isolated sample. **C**: Allele frequencies of two C3 colonies, of which ~50 spores and mycelium were isolated and amplified. **D**: Principal component analysis of different C3 DNA samples and A4, based on allele frequencies of shared SNPs. **E**: MAT locus proportions of different C3 DNA samples and A4. **F**: ddPCR results of other, newly isolated C3 colonies. Note, C3_ddPCR pacbio refers to the WGA amplified version of the DNA shown in lane 1. C3_dd1,2,4,5 and 8 represent unamplified DNA samples for 5 different root culture plates.

The presence of the same MAT loci and near 100% mapping of the A4 Illumina reads to the RirC3 assembly (Additional File 15: Table S1) confirmed the very close relationship between these two isolates. Both strains were harvested as single spores from different parts of the same field in Switzerland and axenic root cultures using *Daucus carota* as host plant were initiated ~20 years ago [39,40]. Many SNPs, even low frequency SNPs, were found to be conserved between A4 and C3, i.e. being variable sites in both, although not necessarily at similar frequencies (Additional File 4). This suggests that they may in fact be clonal strains derived from the same individual.

In most basidiomycete fungi, despite migration of nuclei, exchange of mitochondria does not occur during hyphal anastomosis. Previous studies suggested that anastomoses between closely related AM fungi could lead to exchange of genetically divergent mitochondria [41]. However, only one mitochondrial parental haplotype was found in derived single spore cultures [42]. This has been suggested to occur through an active segregation mechanism by which one mitochondrial haplotype dominated the other. We observed only a single mitochondrial haplotype in C3. Although several low frequency SNPs were found, their number was much lower compared to the SNP frequencies observed in the genomic DNA (Additional File 5). This indicated that the mitochondrial population in this heterogenic strain is also largely homogeneous.

In summary, the characterization of the intragenomic variation showed that there can be substantial variation in allele frequencies between individual cultures of the same strain. To investigate the reasons behind such variation we next looked for signs of inter-nucleus recombination and the segregation of nuclei during sporogenesis, as well as the effect of different host plants on the genetic variation.

Mechanisms behind the observed differences in allele distribution

Potential inter-nucleus recombination

To determine to which extent the observed allele ratios in the genome correlated with the two MAT loci, we sequenced 10 individual nuclei and matched allele variants with their respective MAT locus for each nucleus. Individual nuclei were collected using a fluorescence activated cell sorter and subsequently whole-genome amplified (WGA) before Illumina sequencing [43]. The MAT locus identity of the individual nuclei was determined by PCR analyses (Additional File 10: Fig. S2).

Sequencing reads of individual nuclei were mapped against RirC3 and variants were called in parallel using freebayes (Additional File 6). To avoid confounding effects of putative repetitive sequences or potential mapping/assembly artifacts we only considered SNP's that were outside genomic regions annotated as repeats and homozygous in the single nuclei data based on uniquely mapped reads. Furthermore, SNP's within 500 bp of heterozygous SNPs in single nuclei and/or based on non-paired reads only were omitted. The same analysis was done using Illumina reads of A4 nuclei [19] mapped against RirC3 (Additional File 7). These analyses showed that nuclei clustered together based on MAT locus identity (Fig. 5a,b).

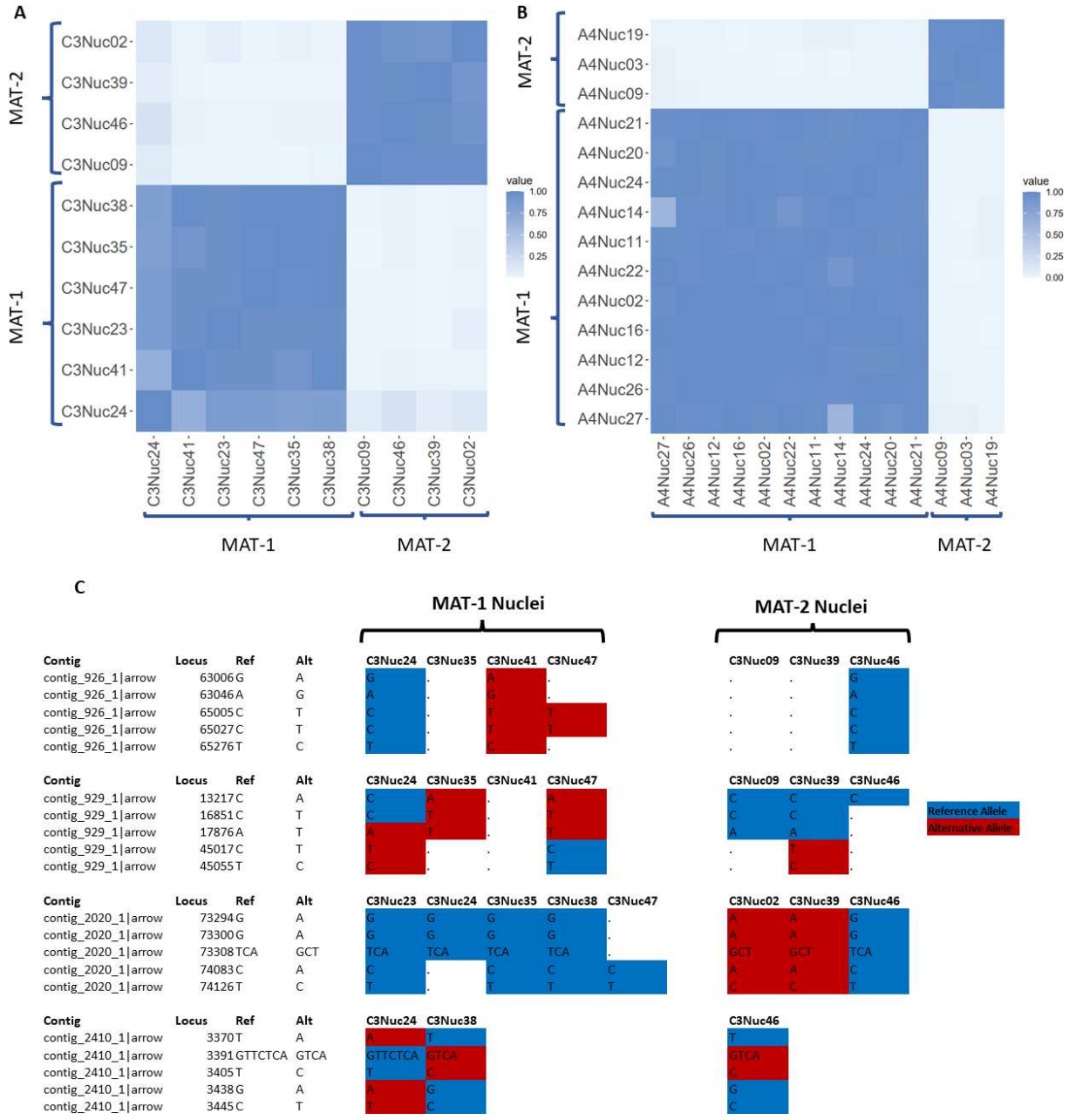


Figure 5. Similarity plots (heat maps) of C3 (**A**) and A4 (**B** [19]) nuclei based on single nucleus sequencing data. Color coding indicates level of relatedness between the nuclei. A sharper contrast between the groups means that the nuclei are more different, while patches of differing colors within the groups indicate similarities to nuclei of the other group (meaning the other MAT locus). Nuclei are grouped based on which MAT locus they contain. **C**: Examples of genotypes of C3 nuclei not consistent with mating type. Indicated in blue the allele typically found in MAT-1 nuclei (ie. the reference allele called in the assembly), in red the allele typically found in MAT-2 nuclei (alternate allele). The MAT

locus identity of the different nuclei is indicated on top. In a true dikaryotic division, all MAT-1 nuclei should have blue alleles, while all MAT-2 nuclei should have red alleles. A complete list of putative recombination sites is given in Additional file 8.

PCA analyses showed that the 6 nuclei containing MAT-1 clustered more closely together with the meta-genomic DNA (C3gDNA), which suggests that this nucleotype contributed the alleles in the assembly (Additional File 11: Fig. S3a). These analyses indicated that most SNPs that were found in the MAT-1 nuclei carried the reference allele called in the assembly, while the MAT-2 nuclei mostly carried the alternative alleles (Fig. 5a; Additional File 6). These analyses further suggested that MAT-2 nuclei are more divergent from each other than the MAT-1 nuclei, which is evident from their clustering less together in the PCA plot (Additional File 11: Fig. S3a). Nuclei with matching MAT loci showed a high level of similarity. Overall, 95% of the SNP's matched the corresponding/expected MAT locus identity, while 5% of the SNP's did not; of the 9947 total SNPs, 503 were represented by both alleles among nuclei with the same MAT allele (Additional File 8). After ignoring contigs where only one SNP was found, 408 SNPs remained covering 89 contigs. Blocks of at least 5 consecutive non-matching SNPs (of 244 total SNPs) were found on 25 contigs.

Previous analyses in A4 suggested that such non-matching sites point to recombination events between nuclei [19], which raised a debate about the validity and amount of recombination sites in the genome [30,31]. To determine whether the number of putative recombination events may be influenced by the quality of the genome assembly, we re-analyzed the single nucleus data from Chen et al. [19] by mapping the A4 single nuclei reads to our RirC3 assembly. All samples mapped at a significantly higher rate and with a higher genome coverage to the RirC3 assembly than to the previous RhiirA4 assembly (Additional File 16). Using identical settings, we found that, compared to the 503 non-matching sites in C3, 97 SNP's (of 5408 total SNP's) did not match their respective MAT locus identity in A4 (Additional File 8). 25 contigs contained multiple non-matching SNPs (total of 62 SNPs),

while only one consecutive block of more than 5 SNPs was found. This suggests a ~ 3 fold difference in the level of SNP's that deviate from a strict (non-recombining) dikaryotic division of nuclei in the two sister strains.

Segregation of nuclei during sporogenesis

Our observed variation in allele frequency distribution between different root culture plates (Fig. 4) raised the suspicion that allele frequencies might be subject to stochastic drift effects. It was previously suggested that varying assortment of genetically different nuclei into newly formed spores can lead to different allele ratios between individual offspring spores [5,7,8,35,44]. This so-called segregation of nuclei could bestow individual single spore offspring lines the ability to differentially affect plant performance. For example, it was suggested that some single spore lines could increase rice growth by a factor of five compared to other lines from the same starting strain [7]. To date, sequence evidence for segregation was only shown for a single repetitive small non-coding locus [35] and its reliability has been questioned [6]. To test for signals of nuclei segregation at spore formation, three single spore lines (root cultures named SS1, SS3 and SS6) were generated from a single ancestral C3 root culture plate. These single spore lines were re-sequenced together with single spores derived from these lines (Fig. 6a). For example, for SS3 one of its single spores was used to generate a second-round single spore line (SS3-1) and a single spore (SS3-1-3) is derived from it. To obtain sufficient material for sequencing all DNA samples were whole genome amplified.

Pattern of allele frequency distribution varied across single spore lines (Fig. 6b-d) and derived individual spores (Additional File 12: Fig. S4), again indicating that nucleotype composition varies between spores within strains. MAT allele ratio was also variable between these samples, showing that MAT locus based nucleotype composition differs between spores (Fig 6e). Differential MAT allele proportions was also supported by ddPCR analyses of the MAT alleles in the same samples (Additional File 13: Fig. S5). PCA analyses based on allele frequencies showed that individual spores varied significantly and no signs of convergence of allele frequencies in next generation spores was observed

(Fig. 6f,g). Intriguingly, DNA isolated from the single spore line SS1 showed an almost exclusively presence of MAT-2 nuclei, with very little MAT-1 nuclei. Nevertheless, a single spore derived from this line (SS1-4) showed a MAT-1:MAT-2 ratio of 8:1, indicating that individual spores can vary widely in their nucleotype composition. Similar, but less extreme, variation was also observed in second- and third-round progeny spores of SS3 and SS6.

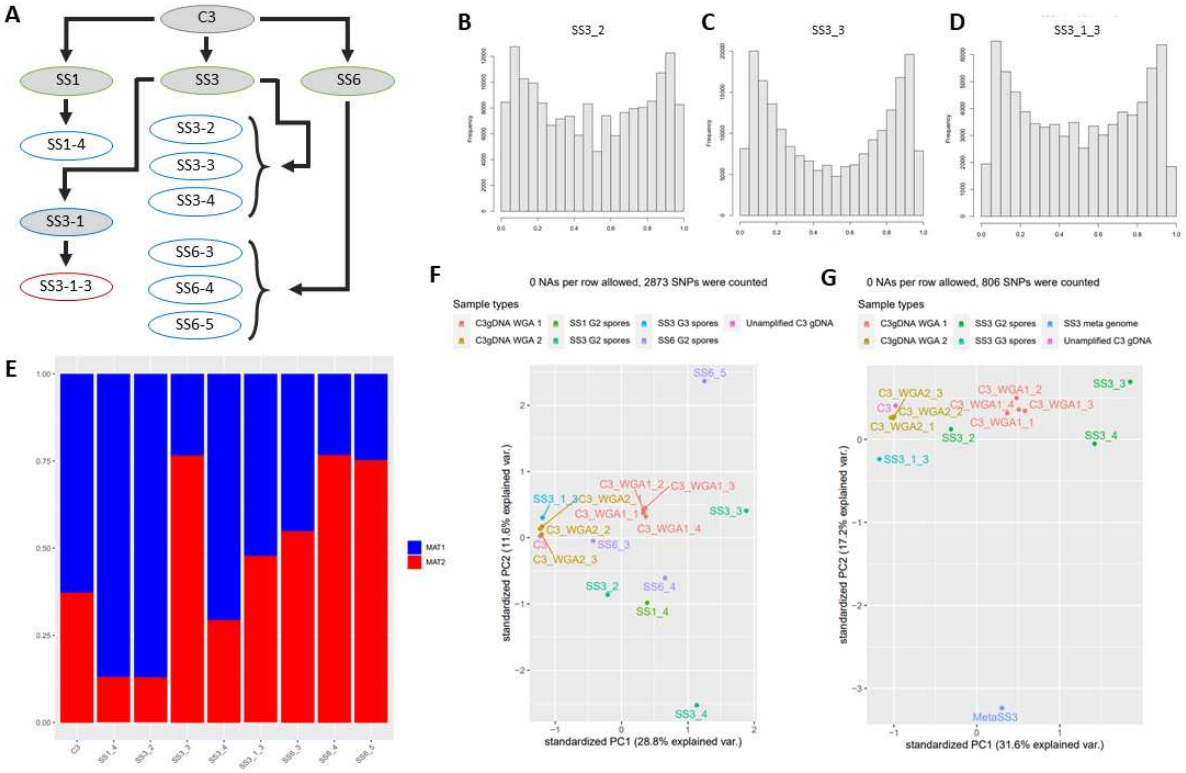
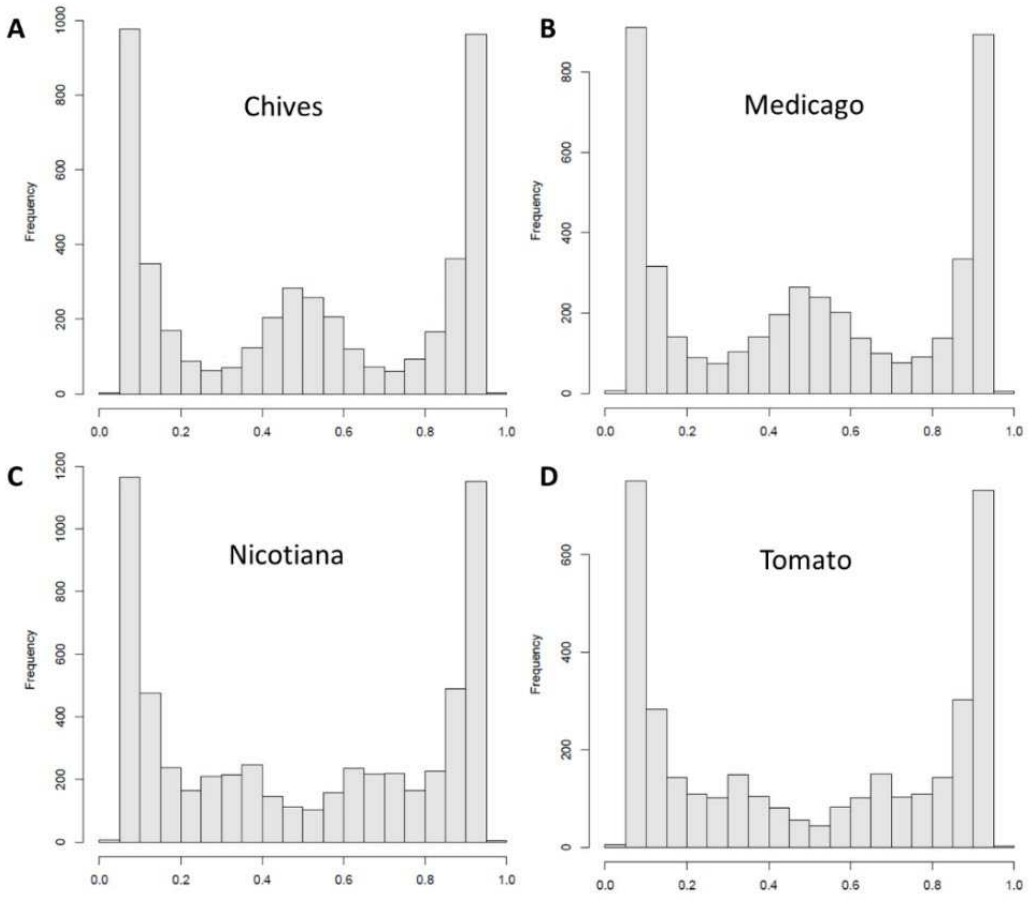


Figure 6. Single spore line variant analysis. **A**: Schematic overview of relatedness of all single spore lines. Lines were created by inoculating *D. carota* root cultures with a single C3 spore. Subsequent generations were made by inoculating a new root culture with a single spore derived from the previous single spore line. Generation (G) number is indicated by color: Black = parental C3 colony, green = G1, blue = G2, red = G3. White circles indicate single spores that were amplified and sequenced, grey circles indicate an established colony producing spores. **B-D**: Allele frequency distributions of several amplified single spores. **E**: MAT loci frequencies of amplified single spores, based on sequencing data. **F**: Principal component analysis of single spores based on allele frequencies of shared SNPs. WGA

samples were included as additional control samples. **G**: Principal component analysis of single spores derived from SS3, including metagenomic DNA from SS3.

Host-dependent differential expression of alleles

While isolates may vary in genotype ratios between the nucleotypes, it is unclear what effect this has on allele expression. To investigate this we performed RNAseq analyses of C3 after colonization of *Medicago truncatula* (Medicago), *Nicotiana benthamiana* (Nicotiana), *Allium schoenoprasum* (Chives) and *Solanum lycopersicum* (Tomato) roots. The spore suspension used for inoculation of the different plants was prepared from a separate host, *D. carota* root culture plates. Strikingly, these analyses revealed two different allele frequency distributions in the fungal mRNA populations depending on host plant identity. In the colonized Medicago and Chive roots the C3 mRNA allele frequencies showed a clear peak at 50%, while in after colonizing Nicotiana and Tomato there was no peak at 50%, and slight allele frequency peaks at ~33% and 67% were observed (Fig. 7a-d). The same observation in three biological replicates of each plant-fungus combination negates batch effects for each inoculation (Additional File 14: Fig. S6). Since all plants were inoculated with the same spore suspension, these data indicated that alleles contributed differently the mRNA pool when colonizing Medicago and Chives, compared to when colonizing Nicotiana or Tomato. Since alleles are distributed over different (haploid) nuclei the genome wide shift in allele frequencies suggests that expression activity varies between nucleotypes. In other words, the contribution of the nuclei to the mRNA pool does not appear to be representative of the nucleotype distribution in the fungus.



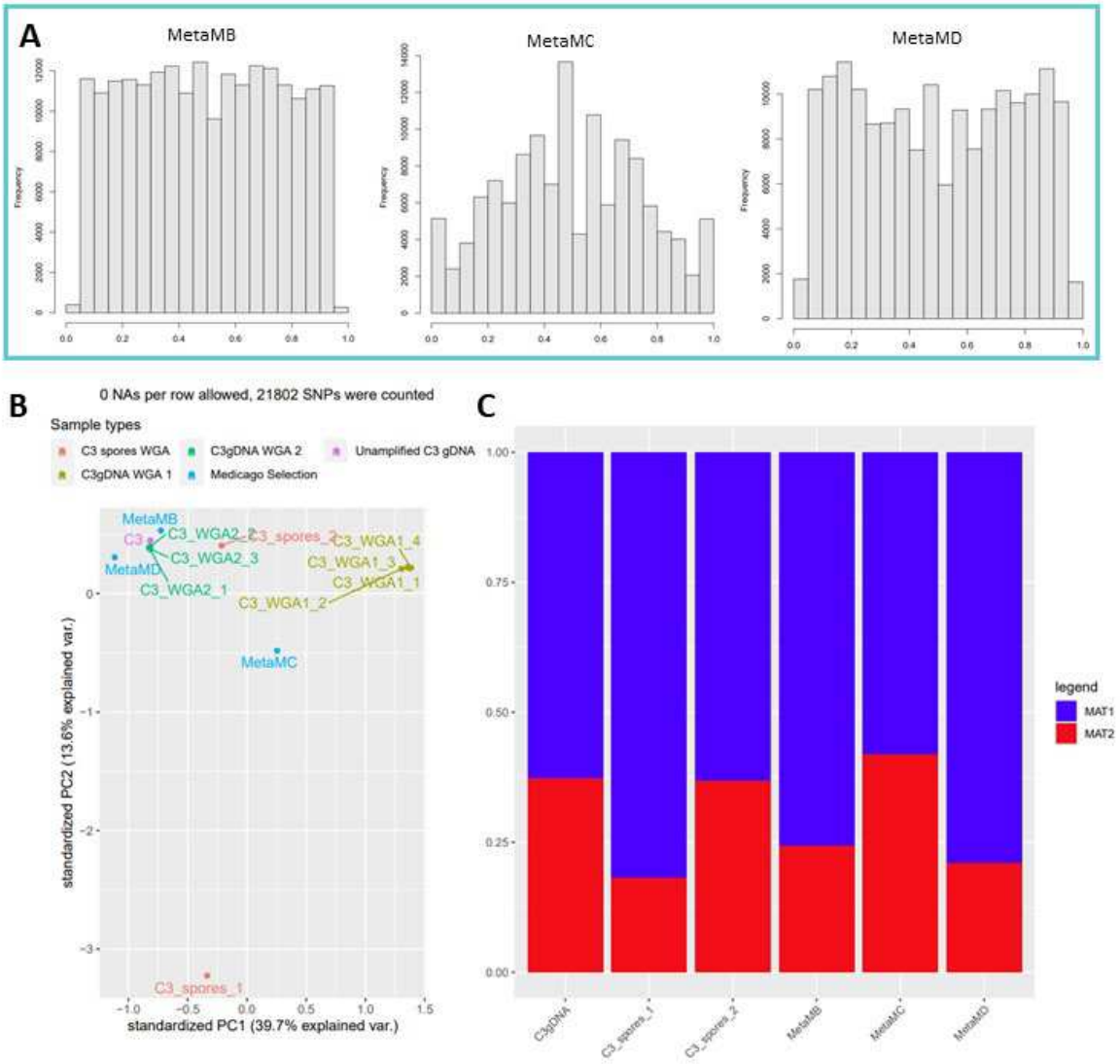
358

359 **Figure 7.** Allele frequency histograms of C3 RNA on different hosts: Chives (**A**), Medicago (**B**),
360 Nicotiana (**C**) and Tomato (**D**). SNPs were filtered on a minimum sequencing depth of 50 reads, with a
361 minimum of 10% alternative alleles.

362

363 Unfortunately, we did not isolate fungal DNA from the corresponding mycorrhized root samples used
364 for RNAseq. This prevented us from testing whether the observed allele frequency distribution in the
365 RNA reflected already a shift in nucleotype ratios at the DNA level due to the different hosts. We have
366 previously hypothesized that genetically different nuclei could have different abilities/efficiencies to
367 interact with distinct plant species [45]. For example, certain nuclei could be more adapted to interact
368 with plant A, whereas other nuclei could be more adapted to interact with plant B. This could in theory
369 lead to a plant effect on the allele frequencies in the offspring when cultured for a longer time on
370 different plant hosts. To determine whether prolonged growth of C3 on Medicago as a host would lead

371 to a consistent shift in of nucleotype ratio's, we performed a selection experiment where we
372 subcultured C3 for four rounds, spanning >2 years, on axenic Medicago root cultures. This resulted in
373 three independent Medicago selection lines that were subsequently sequenced after DNA extraction
374 and whole genome amplification (referred to as MetaMB-D samples in Fig. 8). Unlike the observed 1:1
375 allele frequencies in the mRNA populations, the prolonged co-culturing of C3 with Medicago did not
376 lead to a consistent shift in nucleotype ratios, based on both genome wide allele frequency
377 distributions and MAT allele ratios (Fig. 8a,b; Additional File 12: Fig. S4). PCA analyses based on allele
378 frequencies did not indicate a closer relatedness of Medicago selection lines compare to different
379 batches of *D. carota* root cultures (Fig. 8b). Furthermore, MAT loci frequencies of these lines showed
380 similar variation (Fig. 8c). This contrasts with the findings of [24] for the A4 isolate, where a host effect
381 on nucleotype ratio's was suggested after 1 generation, and argues against a significant effect of host
382 identity on nucleotype distribution in the C3 isolate, at least within the time period investigated.



383

384 **Figure 8:** Selection line variant analysis. **A**: Allele frequency distributions of three Medicago selection

385 lines. **B**: Principal component analysis of C3 and selection lines, based on allele frequencies of shared

386 SNPs. Additional WGA and C3_spores samples were included as additional reference samples. **C**:

387 MAT loci frequency of C3 and selection lines based on read mapping.

388

389

Discussion

Heterokaryosis is common aspect in fungal biology and is hypothesized to play an important role in the ability of fungi to adapt to a continuously changing environment [46]. In case of the extremely coenocytic AM fungi, it was proposed that changes in nucleotype ratio could be adaptive in the colonization of different host plants [24,34,44]. Here, we show that the distribution of genetically divergent nuclei in the *R. irregularis* C3 isolate is highly variable, with seemingly random fluctuations of nucleotype ratio. Large variations in allele frequencies were observed between individual (progeny) spores and single spore lines and even between different root culture plates/batches of the same spore line. Furthermore, no consistent effect of host plant identity on the distribution of nucleotypes was observed after 2 years of subculturing on a different host. Within the time period used, this observation argues against a previously suggested direct effect of host identity on the genetic composition of C3 progeny [24,44,45]. Interestingly, host identity had a reproducible effect on allele-specific expression as observed for C3 grown with three different host plants. These data shed a new light on the genetic organization of AM fungi and further highlight the value of comparing closely related isolates.

The extent of genetic diversity within AM fungal individuals has been highly debated and the current view is that *R. irregularis* strains are either homokaryotic or dikaryotic [6,15,16,18]. Our analyses show that numerous low frequency polymorphisms are not just mere sequencing artefacts, as suggested by Ropars et al. [18], but real components of the genetic variation within AM individuals that is distributed over different nuclei. Furthermore, sequencing of multiple replicate amplifications showed that, although some minor fluctuation in allele frequencies was observed, the whole genome amplification procedure did not introduce significant biases. These results show that the term “dikaryotic” does not fully capture the breath of genetic variation in *R. irregularis* [47], as the coenocytic nature allows for the population of nuclei to accumulate and retain polymorphisms within the nuclear population. This is similar to other fungi, where somatic mutations within an individual lead to polymorphisms that can

be maintained through nuclear selection [48] Intriguingly, multiple low frequency SNPs (occurring between in 10-25% of the reads mapping) were even conserved between the C3 isolate and its presumed clone A4 (Additional File 4). Both strains originated as single spores from different locations in a field in Switzerland and have been individually grown in root cultures for ~20 years [39,40]. Given their high sequence similarity it therefore seems likely that these isolates once originated from the same parental line(s) in the field. Despite all these years of separation, many SNPs have been maintained within the two isolates, even though their allele ratios can vary substantially. In contrast to the variable nucleotype ratio in C3, A4 was reported to show stable nucleotype ratio among different root cultures and individual spores [18,24]. In another *R. irregularis* isolate called SL1, the MAT allele ratio was also found not to be stable across spores and subcultures like in C3 [24]. These data indicate that nuclear dynamics can differ significantly between isolates and even between very closely related strains that originated from the same parental line. What determines this rather different behaviour of nuclei remains to be determined. In line with our data, variable nucleotype ratio's in C3 derived single spore lines were recently also reported based on ddRAD sequencing [49]. In their case the parental C3 line showed mostly 1:1 allele frequencies.

Both C3 and A4 contain two main nucleotypes that can be distinguished based on the sequence diversity of two presumed MAT loci. Single nucleus sequencing revealed at least 503 SNP's in C3 that occurred in different nuclei marked by the same MAT locus. Such SNPs could be the result of somatic mutations and/or point to potential inter-nucleus recombination events between nuclei containing opposing MAT loci. Especially those cases where multiple consecutive SNPs occurred in a single contig and whose allele frequencies in the genome were similar, are strongly suggestive for recombination events. This might be facilitated by the high level of repetitive regions in the genome. However, strikingly the A4 strain showed a much lower amount of putative recombination events when analysed with the same settings using the RirC3 assembly. Although the overall coverage of the A4 single nuclei reads was lower than those of the C3 nuclei (Additional File 15: Table S1), this difference seems unlikely to explain the almost 3 fold difference in non-matching SNP's. This suggests that the level of inter-

nucleus recombination can vary significantly between closely related (sister) strains. Variation in meiotic recombination rates has been reported between and within individuals in other species, i.e. between subsequent measurements or between clones experiencing different environments [50]. A variety of external factors such as temperature, starvation or (biotic) stress, as well as intrinsic genetic or epigenetic mechanisms have been linked to recombination rate plasticity. What underlies the different recombination rates in C3 and A4 and to which extent they are conditional remains to be determined.

The allele frequencies varied wildly among progeny spores as well as compared to their parental lines in a seemingly random fashion. Currently, spore formation is thought to represent the most narrow genetic bottleneck in their life cycle, where the fewest (100-200) nuclei will start a new generation [11]. Single spores that were derived from a previously generated single spore line therefore underwent two genetic bottlenecks compared to the original root culture from which the single spore lines were derived. This could lead to a reduction in genetic variation in subsequent progeny spores. Yet, these second generation single spores were not more similar to each other, but instead varied as much from each other as single spores derived from a different single spore line (Fig. 7). These results illustrate that the genetic composition of a spore is not necessarily representative of the colony that develops from it. At the most extreme we found that the single spore line 1 contained a large majority of MAT-2 nuclei, with very little MAT-1 nuclei. Although we cannot completely rule out that such an extreme ratio is due to the whole genome amplification, we did see large variation in multiple unamplified samples as well. This may suggest that this line would be on its way to a homokaryotic state, however individual progeny spores derived from SS1 again showed completely a completely different ratio.

If segregation of nuclei into developing spores would be a truly random process, modelling suggested that this should lead to a loss of diversity and eventual reversion to a homokaryotic state over time [51,52]. However, the long-term conservation of multiple nucleotypes in C3 indicates that there must

466 be mechanisms that counteract this drift effect. One of these mechanisms may involve continuous
467 nuclear mixing as a result of hyphal fusion/anastomosis, which can occur quite frequently in AMF [53-
468 55]. Modelling showed that such mixing could be sufficient to offset the drift effect [52]. Currently, the
469 dynamics of nuclei are not well understood in AMF. Live cell imaging of hyphae found no evidence for
470 synchronized divisions but showed that nuclei can move in "pulses" in a bi-directional manner,
471 seemingly independent form cytoplasmic streaming [56]. How this movement is regulated or
472 coordinated in different parts of the mycelium is not known, but such pulsed movements could ensure
473 the constant mixing of nuclei facilitating the maintenance of the dikaryotic-like status. Surprisingly, the
474 MAT ratio in single spores of A4 differed from the allele frequencies in mycelial DNA samples [24]. This
475 suggests that there may be a different behavior of the nuclei during spore formation and hyphal
476 growth or in different parts of the mycelium.

477 Fluctuation in nuclear ratios can also be caused by competition between nuclei [57]. However, also in
478 this case, modelling suggested that it would lead to a loss of diversity in favour of the most dominant
479 nucleotype. Therefore, it was proposed that cooperation, or division of labour, between nuclei could
480 lead to the long-term and stable coexistence of distinct genotypes [24]. Also in other fungi, variation
481 in the ratio of nuclear populations have been observed and suggested to be influenced by nuclear
482 selection [12]. The observed stochastic behaviour of nuclei in C3 would argue against a strong
483 interdependence of nucleotypes. In the absence of varying selection pressures inter-nucleus
484 recombination would be expected to reduce diversity in the long term and lead to the fixation of a
485 single nucleotype. This might explain why most of the current AMF cultures appear to be
486 homokaryotic, since the axenic root cultures represent a more or less homogeneous artificial
487 environment with very little variation [58]. In nature, AM fungi will be exposed to continuously
488 changing environments, such as multiple different host plants and soil characteristics with fluctuations
489 in pH, nutrient sources, water availability or other microbes. All these factors may impose different
490 selection pressures which could favour a heterokaryotic state. It would therefore now be interesting

to apply single spore sequencing to spores collected directly from the field to determine the prevalence of dikaryotic-like states, or possibly higher levels of genetic variation.

C3 colonizing Medicago and Chives showed a dominant allele frequency distribution around 50% at the mRNA level for the two MAT nucleotypes, while in the same C3 batch colonizing Nicotiana and Tomato the MAT-1 nuclei appeared to contribute on average ~2x more RNA than the MAT-2 nuclei. This large host-dependent effect on allele frequency distributions at the RNA level indicated that the nucleotypes differ in their transcriptional response. Similar nucleotype-specific expression was recently reported for the multinucleate mushroom *Agaricus bisporus*, which contains two to 25 nuclei of two nuclear types per cell. Widespread transcriptome variation was observed between the two nucleotypes in relation to the development of various *A. bisporus* tissues [59]. This was found to be correlated with differential methylation states, suggesting that epigenetic factors may be important regulators of nucleus-specific expression. An additional level of variation may involve the nucleus-specific expression of distinct ribosomal RNA's. Like DAOM197198, C3 lacked a tandem repeat organization of the 45S rDNA [17]. Eight 45S copies were identified in C3 that showed significant sequence variation and additional polymorphisms were found to be distributed over different nuclei. This may lead to ribosomes with different translational activities in different spores or even different parts of the mycelium [17].

In conclusion, our analyses show that nuclear behavior in *Rhizophagus irregularis* can be highly variable, even between very closely related sister strains. The C3 isolate showed seemingly stochastic nuclear segregation, inter-nucleus genetic variation, significant variation in rDNA variants and nucleotype-specific expression. As the combined output of this genetic variation ultimately determines the effect on plant growth promotion [6], further insight into the nuclear dynamics will be important to understand their distribution and contribution in ecological settings and to exploit their potential as sustainable biofertilizers in agriculture.

Materials and methods

Fungal material

Rhizophagus irregularis isolate C3 was originally isolated from Tänikon, Switzerland. The fungus was propagated on *Agrobacterium rhizogenes*-transformed *Daucus carota* root cultures on M medium [61,62].

Single spore lines were generated by placing a single C3 spore next to a fresh *D. carota* root culture. Spores were selected from spore clusters from the same source plate, and single spore lines were named after their respective cluster.

Medicago selection lines (MedSel) were made by inoculating *Medicago truncatula* (Jemalong A17) root cultures with ~50 C3 spores. When these cultures produced enough spores, these spores moved to fresh *M. truncatula* root cultures to start a new round. Three of these subsequent transfers were made. For DNA sequencing, ~50 spores were isolated from the M medium and crushed in 2µL DNA free mQ water. Total genomic DNA was then amplified using the Repli-G WGA kit (Qiagen).

DNA isolation for genome assembly

Four square plates, six round plates and four split plates containing fully C3 mycorrhized *D. carota* root cultures were harvested and pooled. Upon harvesting the fungal material, roots were removed from root culture plates with pliers and scalpel, after which the medium was liquidized by adding 1/2 volumes 100mM Citrate buffer (40mM sodium citrate dihydrate, 60mM citric acid, pH = 6.5) to each volume of M medium and shaking at RT for at least 30 minutes. The dissolved medium was then poured into an empty square petri dish, from which the mycelium and spores were collected with a sterile disposable inoculator loop, while taking care to avoid any pieces of the root culture. Collected spores and mycelium were washed in sterile mQ, collected in a 2mL Eppendorf tube and centrifuged at 5000rpm. As much water as possible was removed from the tube, after which the sample was weighed and flash-frozen in liquid nitrogen.

Samples were thoroughly (>20x 20s) pulverized with a metal bead in a TissueLyser LT (Qiagen). All materials were kept at minimal temperatures to avoid thawing of the sample. For the isolation of high molecular weight genomic DNA, a protocol from Fauchery *et al.* [63] was adapted. The lysis buffer was made of five stock solutions (Additional File 16: Table S2) that were combined shortly before the isolation. 1.5mL of the lysis buffer was added to the frozen fungal material. The sample was mixed by gently shaking until the sample was completely suspended in the buffer. Lysis was performed at 65 ^{0}C for 30 minutes, gently shaking every 10 minutes. The lysis was stopped by adding 492µL 5M KAc (pH 7.5) and gently inverting. The sample was incubated on ice for 30 minutes and centrifuged at 5000g at 4C for 20 minutes. The supernatant was transferred to a 15mL Falcon tube, and cleaned by adding 1 volume chloroform–isoamyl alcohol (24:1 v/v), gently but thoroughly shaking until completely mixed, and pipetting the upper layer to a fresh tube. This step was repeated twice to remove all residual proteins. 10µL RNAse A (10mg/ml) was added and the sample was incubated at 37 ^{0}C for 1 hour. Next, 20µL 3M NaAc (pH = 5.2) was added, the sample was mixed, and then precipitated by adding 1 volume of isopropanol. The sample was incubated at RT for 15minutes before centrifuging at 4 ^{0}C for 30 minutes at max speed. The supernatant was discarded and the pellet was washed with ice cold 70% ethanol. The sample was then dried at RT and resuspended in 55µL 20mM Tris-HCl at 65% for 30 minutes. 5µL of the solution was diluted 4x for quality control, the rest was immediately stored at -70 ^{0}C . Yield was measured by Qubit 2.0 fluorometer via the Qubit dsDNA HS Assay (Life Technologies) and DNA integrity was checked on 0.8% agarose gel. 900ng of high molecular weight genomic DNA was collected for PacBio sequencing.

PacBio assembly

PacBio SMRT subreads were generated at GenomeScan B.V. (Leiden, The Netherlands) The subreads were assembled using Flye (2.7.1-b1590) [64] with the following command: flye --pacbio-raw C3_PacBio_subreads.fastq.gz -g 156m --out-dir C3_assembly --threads 30. Duplicated regions were removed with purge_dups [65]. Genome polishing was performed in two steps: first using the PacBio

subreads using Arrow (Pacific Biosciences), then with Illumina reads of C3 using two iterations of Racon (v1.4.13) [66] C3 Illumina reads were produced by sequencing 300ng of C3 genomic DNA, isolated from the same cultures as the PacBio sample, at NovoGene. Genome completeness was assessed with BUSCO [67], using database fungi_db10. Repeats were modeled *de novo* with RepeatModeler and subsequently masked with RepeatMasker (v. open-4.0.9) [68]. The genome was annotated using Funannotate (v1. 6.0) [37], using predicted gene models and C3 RNAseq reads from C3 grown on multiple hosts (see RNAseq section). The mitochondrial genome was found by blasting RhiirA4 mitochondrial markers [18] against the raw RirC3 assembly (before purge_dups). All markers were found on a single contig covering the entire predicted mitochondrial genome. Ribosomal DNA copies were found by blasting Rir17 rDNA sequences in the RirC3 assembly. Contigs of the assembly were visualized with Circos [69].

Variant calling

Illumina reads of C3 were mapped against the RirC3 assembly using Hisat2 [70], and sorted with samtools sort -o mappedreads.bam. Variant calling was performed on mapped reads using Freebayes (v1.3.2) [71], setting ploidy level to 1 with the pooled-discrete -J option. Variants were filtered using bcftools filter (v 1.10.2) [72], selecting SNPs within the 25 percentile above and below average mapping depth, and reference and alterative allele observation of at least 10. Allele frequency distributions were plotted in R (v 4.0.3) using the hist() command. Principal component analysis was performed by merging vcf files with bcftools merge, and creating a dataframe of all allele frequencies in R. Next, only SNPs with coverage in all samples were selected. Principal component analysis was performed using prcomp(df, center = TRUE, scale. = TRUE), and plotted using ggbiplot().

Single nuclei isolation and sequencing

Spores of C3 were suspended in 1xPBS buffer (pH 7.4) and crushed with a pestle. Nuclei were selected by fluorescent associated cell sorting (FACS) [43] and whole genome amplified (WGA) through MDA using Phi29 polymerase. An 80x dilution of the reactions was used for genotyping. The remaining

reaction mixture was purified using ethanol precipitation and dissolved in 30µL 10mM Tris-HCl. Genotyping was done using primers targeting the ITS region (AM1 + NS31) [73,74] or MAT loci (Forward: ACTATCTGACTTGCTATTGTTGA, Reverse: CAGGGCCTGCATCGGATTA). Ten of the nuclei were sent for Illumina sequencing (NovoGene, Hong Kong). Reads were mapped against RirC3 using HiSat2 [70] with standard settings, and variant calling was performed with freebayes (p = 1). Variants were selected by first intersecting the vcf file with the filtered RirC3 gDNA vcf file, only containing non-repeated regions. Heterozygous SNPs were found and filtered out using the Awk utility in Unix, removing heterozygous SNPs which were defined as having a frequency above 10% (RO/AO > 0.1 || AO/RO > 0.1) in any of the nuclei. Similarity plots were made in R using ggplot(). Individual nuclear genomes were assembled using Spades [43,75].

To find potential recombination sites, vcf files were grouped based on MAT identity. To avoid potential only uniquely mapping reads were considered. SNP's within 500 bp upstream and downstream of heterozygous loci in and/or based on non-paired reads only, or with a coverage below 10, were ignored. If genotypes of any SNP in the genome were linked to MAT identity of the nuclei, all nuclei sharing a MAT locus should have the same genotype on that SNP. Therefore, any SNP where both the reference and alternative allele were found in nuclei with the same MAT locus, was considered as a potential recombination event.

Single spore amplification and analysis

Single spores were isolated by excising M medium containing spores from root cultures, and subsequently dissolving them in citrate buffer. The spores were thoroughly rinsed with sterile mQ water and collected in 2µL of mQ in 200µL PCR strips. Spores were manually crushed using pipette tips of which the tips were briefly melted in an open flame, to create a "pestle". After crushing, the samples were flash-frozen in liquid nitrogen and incubated at 95 ^{0}C for 10 minutes to further lyse the nuclei. WGA was performed using the Repli-G Single Cell kit (Qiagen) following manufacturer's instructions.

The samples were purified by ethanol precipitation and dissolved in 30µL 10mM Tris-HCl buffer. Samples were sent for Illumina sequencing in NovoGene (Hong Kong).

Plant inoculation for RNAseq

C3 inoculum

Spores of C3 were released from root cultures by disrupting the root cultures in a blender with 2x volume water and filtering with a 40µm mesh to capture the spores and mycorrhized root fragments. Spore suspensions were stored at 4 ^{0}C.

Nicotiana

Nicotiana benthamiana seeds were sterilized in 20% bleach solution for 12 minutes, thoroughly washed with sterile water and germinated on water agar with a filter for 72h at RT, in 16/8 light dark cycle. Pots (9x11x11cm) with 2:1 sterilized clay:silver sand mix were prepared, and ~200 C3 spores were added in the middle of the pot, ~4cm below surface. The middle of the surface was covered with a small amount of 1:1 vermiculite/silver sand mix, to act as a more stable soil to plant the small seedling. After germination, seedlings were planted with a fine brush in a 1mm hole in the vermiculite/silver sand mix. To increase initial growth rate, 2mL of high phosphate half-strength Hoagland solution [76] (1 mM K2PO4) was added to the seedlings. The pots were covered with plastic foil for the first week to maintain soil humidity, and plants were watered twice a week with adjusted low phosphate half-strength Hoagland solution (50µM K2PO4). Plants were grown at 25 ^{0}C in a 16/8 light dark cycle. Mycorrhized roots were harvested after six weeks by gently submerging the pot in water, removing soil and clay from the roots under water and rinsing carefully with tap water.

Medicago

Medicago truncatula Jemalong A17 seeds were scarified in 96% sulfuric acid for 10 minutes, thoroughly rinsed with water, sterilized with 50% bleach for 10 minutes and washed again with sterilized water. Seeds were then incubated on a water agar plate with filter at 4 ^{0}C in the dark, and then incubated at

638 21 ^{0}C in the dark. Pots (9x11x11cm) with 1:1 sterilized clay:silver sand mix were prepared, and ~200 C3
639 spores were added in the middle of the pot, ~4cm below the surface. After germinating, seedlings were
640 planted in the pots. Plants were grown at 21 ^{0}C in a 16/8 light dark cycle, and were watered with
641 adjusted half-strength Hoagland solution (20µM K2PO4). Mycorrhized roots were harvested after six
642 weeks and gently washed with tap water.

643 *Chives*

644 *Allium schoenoprasum* seeds were soaked in mQ at 4 ^{0}C for 8h, disinfected with 20% bleach for 12
645 minutes and thoroughly rinsed with sterile water. Seeds were germinated on agar plates with a filter
646 for 48h at 21 ^{0}C (16/8 light dark cycle). Pots (9x11x11cm) with 1:1 sterilized clay:silver sand mix were
647 prepared, and ~200 C3 spores were added in the middle of the pot, ~4cm below the surface. After
648 germinating, seedlings were planted in the pots. Plants were grown at 21 ^{0}C (16/8 light dark cycle), and
649 were watered with adjusted half-strength Hoagland solution (20µM K2PO4). Mycorrhized roots were
650 harvested after six weeks and gently washed with tap water.

651 *Tomato*

652 *Solanum lycopersicum* (MoneyMaker) seeds were soaked in 0.03M HCl for 6h and sterilized in 50%
653 bleach for 5 minutes. Seeds were then germinated at 25 ^{0}C for 72h in a 16/8 light dark cycle. Large pots
654 (18x11x11cm) were filled with a 2:1:1 clay:silver sand:vermiculite mix. ~200 C3 spores were placed
655 ~8cm below surface. After germination, seedlings were planted and watered with adjusted half-
656 strength Hoagland solution (50µM K2PO4). Mycorrhized roots were harvested after eight weeks by
657 gently submerging the pot in water, removing vermiculite and clay from the roots under water and
658 rinsing carefully with tap water.

659 **RNA isolation and sequencing**

660 RNA from colonized roots was isolated by flash-freezing colonized roots and destroying the tissue with
661 a cold mortar and pestle. RNA isolation was performed using the RNeasy Mini kit (Qiagen), according

to manufacturer's instructions including an on-column RNAse free DNAse (Qiagen) treatment. Three biological replicates of each treatment were sent for Illumina sequencing (BGI, Denmark). RNAseq reads were mapped to the RirC3 assembly with Hisat2 using the --dta option. Variant calling was performed with Freebayes as described above, variants were filtered based on a minimal sequencing depth of 50 with at least 5% alternative alleles.

Digital droplet PCR

Digital droplet PCR was performed using 80 ng/µl each of MAT-1 and MAT-2 specific primers [24] and QX200™ ddPCR™ EvaGreen Supermix (BioRad) in a total volume of 20 µl. For unamplified (meta-)samples 2 µl of a 1:10 dilution was used per reaction. For WGA amplified samples, 2 µl of a 1:100 dilution was used as template. The PCR mix was suspended in oil for EvaGreen using the QX200 Droplet Generator (Biorad), following manufacturer's instructions. PCR was performed for 40 cycles, annealing and elongation at 58 0C. Subsequently, the absolute number of positive droplets was counted using a QX200 Droplet Reader and analysed via QuantaSoft Software (BioRad).

Availability of data and materials

All C3 sequencing data and genome assembly generated in this work are available from Genbank under BioProject ID PRJNA747641 and SRR15179489 - SRR1517934. Sequencing data for A4 were retrieved from BioProject ID PRJNA299206 and PRJNA477348.

Supplementary Information

Additional file 1: RirC3 BUSCO output and comparison.

Additional file 2: Additional C3 45S rDNA polymorphisms (.vcf file).

Additional file 3: Allele variant (SNP) list based on C3 gDNA reads (.vcf file).

685 **Additional file 4**: Allele variant (SNP) comparison of C3 and A4 (.vcf file).

686 **Additional file 5**: C3 mitochondrial DNA variants (.vcf file).

687 **Additional file 6**: Allele variant (SNP) table of 10 single nuclei from C3 (.vcf file).

688 **Additional file 7**: Allele variant (SNP) table of 14 single nuclei from A4 (reads from Chen et al., 2018)
689 (.vcf file).

690 **Additional file 8**: Potential inter-nucleus recombination events in C3 and A4.

691 **Additional file 9**: Fig. S1. Allele frequency distribution of replicate, independently WGA-amplified, C3
692 gDNA samples; corresponding to main Fig. 4A,B. Two replicates for C3-gDNA2 and 3 replicates for C3-
693 gDNA1 (used for genome assembly).

694 **Additional file 10**: Fig. S2. PCR analysis of MAT locus identity in C3 single nuclei. The upper band
695 corresponds to MAT-1, the lower band to MAT-2.

696 **Additional file 11**: Fig. S3. Principal component analysis of C3 single nuclei (**A**) and A4 single nuclei (**B**)
697 based on allele frequencies when mapped to the RirC3 assembly. The MAT locus identity of the
698 individual nuclei is indicated by color: red = MAT-1, blue = MAT-2.

699 **Additional file 12**: Fig. S4. Allele frequency analysis of (WGA amplified) C3 single spores derived from
700 single spore lines.

701 **Additional file 13**: Fig. S5. MAT ratio based on digital droplet PCR of different root culture
702 batches/lines. C3 pac bio refers to DNA sample C3 gDNA2 used for genome assembly, either
703 unamplified or WGA amplified (Amp1). C3 old refers to the independent DNA sample C3 gDNA1.
704 C3_dd1 and _dd2 refer to WGA amplified DNA from two additional independent C3 carrot root
705 culture batches. C3 meta1 and meta2 refer to DNA extracted and WGA amplified from groups of 50
706 spores from two different root culture plates. SS1 refers to non-amplified DNA from single spore line

1. MB, MC and MD refer to non-amplified DNA from three Medicago selection lines. SS1_4, SS3_2, SS3_3 and SS6_3 refer to DNA samples from 2nd generation single spore lines.

Additional file 14: Fig. S6. Allele frequency analysis based on RNAseq data from two additional biological replicate samples of C3 colonizing Chives, Medicago, Nicotiana and Tomato; corresponding to main Fig. 7.

Additional file 15: Table S1. Comparison of mapping rate and genome coverage of A4 gDNA and single nuclei data (from [19]), mapped against the RhiiA4 assembly [18] and RirC3.

Additional file 16: Table S2. Composition of the lysis buffer mix used for gDNA extraction, used for PacBio sequencing.

Funding

JvC is supported by the Dutch Experimental Plant Sciences research school (EPS Project 3184319610). BA is supported by the Netherlands Organisation of Scientific Research (NWO: ALWGR.2017.010). RW is supported by the Netherlands Organisation of Scientific Research (NWO: grant ALWGR.2015.9). AR is supported by the European Research Council (ERC (678792).

Acknowledgments

The original C3 carrot root culture was kindly provided by prof. Ian Sanders (University of Lausanne, Switzerland). We further like to thank Guido Hooiveld and Mara van Trijp (Human Nutrition and Health, Wageningen University & Research) for assistance with the ddPCR. Nuclei sorting and their whole genome amplification was done at the SciLifeLab Microbial Single Cell Genomics Facility at Uppsala University.

Author contributions

JvC and EL conceived the study and designed the experiments. JvC, CB and EL performed experiments. JvC, RW and BA performed computational analyses. JA and ZP supported the Illumina sequencing. JvC and EL wrote the paper, with input of BA, RW, AR and TB. All authors read and approved the final manuscript.

The authors declare no conflict of interest.

738 **References**

739 1 Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N,
740 Grigoriev I, Gryganskyi A, James TY. A phylum-level phylogenetic classification of
741 zygomycete fungi based on genome-scale data. Mycologia. 2016 Sep 1;108(5):1028-46.
742 2 Luginbuehl LH, Oldroyd GE. Understanding the Arbuscule at the Heart of
743 Endomycorrhizal Symbioses in Plants. Curr Biol. 2017;27(17):R952-R963. doi:
744 10.1016/j.cub.2017.06.042.
745 3 Redecker D, Morton JB, Bruns TD. Ancestral lineages of arbuscular mycorrhizal fungi
746 (Glomales). Molecular phylogenetics and evolution. 2000 Feb 1;14(2):276-84.
747 4 Bruns TD, Corradi N, Redecker D, Taylor JW, Öpik M. Glomeromycotina: what is a species
748 and why should we care?. New Phytologist. 2018 Dec;220(4):963-7.
749 5 Sanders IR, Croll D. Arbuscular mycorrhiza: the challenge to understand the genetics of
750 the fungal partner. Annual review of genetics. 2010 Dec 1;44:271-92.
751 6 Kokkoris V, Stefani F, Dalpé Y, Dettman J, Corradi N. Nuclear dynamics in the arbuscular
752 mycorrhizal fungi. Trends in Plant Science. 2020 Aug 1;25(8):765-78.
753 7 Angelard C, Colard A, Niculita-Hirzel H, Croll D, Sanders IR. Segregation in a mycorrhizal
754 fungus alters rice growth and symbiosis-specific gene transcription. Current Biology. 2010
755 Jul 13;20(13):1216-21.
756 8 Ehinger MO, Croll D, Koch AM, Sanders IR. Significant genetic and phenotypic changes
757 arising from clonal growth of a single spore of an arbuscular mycorrhizal fungus over
758 multiple generations. New Phytologist. 2012 Nov;196(3):853-61.
759 9 Balestrini R, Bianciotto V, Bonfante-Fasolo P. Nuclear architecture and DNA location in
760 two VAM fungi. Mycorrhiza. 1992 Jun 1;1(3):105-12.
761 10 Jany JL, Pawlowska TE. Multinucleate spores contribute to evolutionary longevity of
762 asexual glomeromycota. The American Naturalist. 2010 Apr;175(4):424-35.

763 11 Marleau J, Dalpé Y, St-Arnaud M, Hijri M. Spore development and nuclear inheritance in
764 arbuscular mycorrhizal fungi. BMC evolutionary biology. 2011 Dec;11(1):1-1.
765 12 Roper M, Simonin A, Hickey PC, Leeder A, Glass NL. Nuclear dynamics in a fungal chimera.
766 Proceedings of the National Academy of Sciences. 2013 Aug 6;110(32):12875-80.
767 13 Roberts SE, Gladfelter AS. Nuclear autonomy in multinucleate fungi. Current opinion in
768 microbiology. 2015 Dec 1;28:60-5.
769 14 Martin F, Aerts A, Ahrén D, Brun A, Danchin EG, Duchaussoy F, Gibon J, Kohler A,
770 Lindquist E, Pereda V, Salamov A. The genome of Laccaria bicolor provides insights into
771 mycorrhizal symbiosis. Nature. 2008 Mar;452(7183):88-92.
772 15 Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing
773 N, dit Frey NF, Gianinazzi-Pearson V, Gilbert LB. Genome of an arbuscular mycorrhizal
774 fungus provides insight into the oldest plant symbiosis. Proceedings of the National
775 Academy of Sciences. 2013 Dec 10;110(50):20117-22.
776 16 Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DG, Mu D, Pang E, Cao H, Cha H, Lin T, Zhou
777 Q. Single nucleus genome sequencing reveals high similarity among nuclei of an
778 endomycorrhizal fungus. PLoS genetics. 2014 Jan 9;10(1):e1004078.
779 17 Maeda T, Kobayashi Y, Kameoka H, Okuma N, Takeda N, Yamaguchi K, Bino T, Shigenobu
780 S, Kawaguchi M. Evidence of non-tandemly repeated rDNAs and their intragenomic
781 heterogeneity in Rhizophagus irregularis. Communications biology. 2018 Jul 10;1(1):1-3.
782 18 Ropars J, Toro KS, Noel J, Pelin A, Charron P, Farinelli L, Marton T, Krüger M, Fuchs J,
783 Brachmann A, Corradi N. Evidence for the sexual origin of heterokaryosis in arbuscular
784 mycorrhizal fungi. Nature Microbiology. 2016 Mar 21;1(6):1-9.
785 19 Chen EC, Mathieu S, Hoffrichter A, Sedzielewska-Toro K, Peart M, Pelin A, Ndikumana S,
786 Ropars J, Dreissig S, Fuchs J, Brachmann A. Single nucleus sequencing reveals evidence of
787 inter-nucleus recombination in arbuscular mycorrhizal fungi. Elife. 2018 Dec 5;7:e39813.

20 Kobayashi Y, Maeda T, Yamaguchi K, Kameoka H, Tanaka S, Ezawa T, Shigenobu S, Kawaguchi M. The genome of Rhizophagus clarus HR1 reveals a common genetic basis for auxotrophy among arbuscular mycorrhizal fungi. BMC genomics. 2018 Dec;19(1):1-1.

21 Venice F, Ghignone S, Salvioli di Fossalunga A, Amselem J, Novero M, Xianan X, Sędzielewska Toro K, Morin E, Lipzen A, Grigoriev IV, Henrissat B. At the nexus of three kingdoms: the genome of the mycorrhizal fungus Gigaspora margarita provides insights into plant, endobacterial and fungal interactions. Environmental microbiology. 2020 Jan;22(1):122-41.

22 Halary S, Malik SB, Lildhar L, Slamovits CH, Hijri M, Corradi N. Conserved meiotic machinery in Glomus spp., a putatively ancient asexual fungal lineage. Genome biology and evolution. 2011 Jan 1;3:950-8.

23 Tisserant E, Kohler A, Dozolme-Seddas P, Balestrini R, Benabdellah K, Colard A, Croll D, Da Silva C, Gomez SK, Koul R, Ferrol N. The transcriptome of the arbuscular mycorrhizal fungus Glomus intraradices (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. New Phytologist. 2012 Feb;193(3):755-69.

24 Kokkoris V, Chagnon PL, Yildirir G, Clarke K, Goh D, MacLean AM, Dettman J, Stefani F, Corradi N. Host identity influences nuclear dynamics in arbuscular mycorrhizal fungi. Current Biology. 2021 Apr 12;31(7):1531-8.

25 Jinks JL. Heterokaryosis: a system of adaptation in wild fungi. Proceedings of the Royal Society of London. Series B-Biological Sciences. 1952 Aug 27;140(898):83-99.

26 Rep M, Kistler HC. The genomic organization of plant pathogenicity in Fusarium species. Current opinion in plant biology. 2010 Aug 1;13(4):420-6.

27 Roper M, Ellison C, Taylor JW, Glass NL. Nuclear and genome dynamics in multinucleate ascomycete fungi. Current biology. 2011 Sep 27;21(18):R786-93.

28 Croll D, Giovannetti M, Koch AM, Sbrana C, Ehinger M, Lammers PJ, Sanders IR. Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus Glomus intraradices. New Phytologist. 2009 Mar;181(4):924-37.

29 Novais CB, Pepe A, Siqueira JO, Giovannetti M, Sbrana C. Compatibility and incompatibility in hyphal anastomosis of arbuscular mycorrhizal fungi. Scientia Agricola. 2017 Sep;74:411-6.

30 Chen EC, Mathieu S, Sedzielewska-Toro K, Peart M, Pelin A, Ndikumana S, Ropars J, Dreissig S, Fuchs J, Brachmann A, Corradi N. Correction: Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. Elife. 2019 Mar 21;8:e46860.

31 Auxier B, Bazzicalupo A. Comment on'Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi'. Elife. 2019 Oct 25;8:e47301.

32 Forche A, Alby K, Schaefer D, Johnson AD, Berman J, Bennett RJ. The parasexual cycle in Candida albicans provides an alternative pathway to meiosis for the formation of recombinant strains. PLoS biology. 2008 May;6(5):e110.

33 Croll D, Wille L, Gamper HA, Mathimaran N, Lammers PJ, Corradi N, Sanders IR. 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. 2008 May;178(3):672-87.

34 Wyss T, Masclaux FG, Rosikiewicz P, Pagni M, Sanders IR. Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus Rhizophagus irregularis. The ISME journal. 2016 Oct;10(10):2514-26.

35 Masclaux FG, Wyss T, Pagni M, Rosikiewicz P, Sanders IR. Investigating unexplained genetic variation and its expression in the arbuscular mycorrhizal fungus *Rhizophagus*

irregularis: A comparison of whole genome and RAD sequencing data. PloS one. 2019 Dec 27;14(12):e0226497.

36 Miyauchi S, Kiss E, Kuo A, Drula E, Kohler A, Sánchez-García M, Morin E, Andreopoulos B, Barry KW, Bonito G, Buée M. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. Nature communications. 2020 Oct 12;11(1):1-7.

37 Palmer J, Nextgenusfs SJ. Funannotate: Funannotate v1. 6.0.

38 Viruel J, Conejero M, Hidalgo O, Pokorny L, Powell RF, Forest F, Kantar MB, Soto Gomez M, Graham SW, Gravendeel B, Wilkin P. A target capture-based method to estimate ploidy from herbarium specimens. Frontiers in plant science. 2019 Jul 24;10:937.

39 Jansa J, Mozafar A, Anken T, Ruh R, Sanders I, Frossard E. Diversity and structure of AMF communities as affected by tillage in a temperate soil. Mycorrhiza. 2002 Oct;12(5):225-34.

40 Koch AM, Kuhn G, Fontanillas P, Fumagalli L, Goudet J, Sanders IR. High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. Proceedings of the National Academy of Sciences. 2004 Feb 24;101(8):2369-74.

41 de la Providencia IE, Nadimi M, Beaudet D, Rodriguez Morales G, Hijri M. Detection of a transient mitochondrial DNA heteroplasmy in the progeny of crossed genetically divergent isolates of arbuscular mycorrhizal fungi. New Phytologist. 2013 Oct;200(1):211-21.

42 Daubois L, Beaudet D, Hijri M, de la Providencia I. Independent mitochondrial and nuclear exchanges arising in Rhizophagus irregularis crossed-isolates support the presence of a mitochondrial segregation mechanism. BmC microbiology. 2016 Dec;16(1):1-2.

43 Montoliu-Nerin M, Sánchez-García M, Bergin C, Grabherr M, Ellis B, Kutschera VE, Kierczak M, Johannesson H, Rosling A. Building de novo reference genome assemblies of

complex eukaryotic microorganisms from single nuclei. Scientific reports. 2020 Jan 28;10(1):1-0.

44 Angelard C, Tanner CJ, Fontanillas P, Niculita-Hirzel H, Masclaux F, Sanders IR. Rapid genotypic change and plasticity in arbuscular mycorrhizal fungi is caused by a host shift and enhanced by segregation. ISME J. 2014;8(2):284-94. doi: 10.1038/ismej.2013.154.

45 Limpens E, Geurts R. Plant-driven genome selection of arbuscular mycorrhizal fungi. Molecular plant pathology. 2014 Aug;15(6):531.

46 Strom NB, Bushley KE. Two genomes are better than one: history, genetics, and biotechnological applications of fungal heterokaryons. Fungal Biology and Biotechnology. 2016 Dec;3(1):1-4.

47 Masclaux FG, Wyss T, Pagni M, Rosikiewicz P, Sanders IR. Investigating unexplained genetic variation and its expression in the arbuscular mycorrhizal fungus *Rhizophagus irregularis*: A comparison of whole genome and RAD sequencing data. PLoS ONE 2019;14(12):e0226497. DOI: 10.1371/journal.pone.0226497

48 Grum-Grzhimaylo AA, Bastiaans E, van den Heuvel J, Berenguer Millanes C , Debets AJM, Aanen DK. Somatic deficiency causes reproductive parasitism in a fungus. Nat Commun 202;12(1):783. doi: 10.1038/s41467-021-21050-5.

49 Robbins C, Cruz Corella J, Aletti C, Seiler R, Mateus ID, Lee SJ, Masclaux FG, Sanders IR. Generation of disproportionate nuclear genotype proportions in Rhizophagus irregularis progeny causes allelic imbalance in gene transcription. New Phytologist. 2021 Jun 4.

50 Stapley J, Feulner PG, Johnston SE, Santure AW, Smadja CM. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017 Dec 19;372(1736):20160455.

51 Bever JD, Wang M. Hyphal fusion and multigenomic structure. Nature. 2005 Jan;433(7022):E3-4.

52 Bever JD, Kang HJ, Kaonongbua W, Wang M. Genomic organization and mechanisms of inheritance in arbuscular mycorrhizal fungi: contrasting the evidence and implications of current theories. InMycorrhiza 2008 (pp. 135-148). Springer, Berlin, Heidelberg.

53 Giovannetti M, Azzolini D, Citernesi AS. Anastomosis formation and nuclear and protoplasmic exchange in arbuscular mycorrhizal fungi. Applied and environmental microbiology. 1999 Dec 1;65(12):5571-5.

54 Giovannetti M, Fortuna P, Citernesi AS, Morini S, Nuti MP. The occurrence of anastomosis formation and nuclear exchange in intact arbuscular mycorrhizal networks. New Phytologist. 2001 Sep;151(3):717-24.

55 De La Providencia IE, De Souza FA, Fernández F, Delmas NS, Declerck S. Arbuscular mycorrhizal fungi reveal distinct patterns of anastomosis formation and hyphal healing mechanisms between different phylogenic groups. New Phytologist. 2005 Jan;165(1):261-71.

56 Bago B, Zipfel W, Williams RM, Piché Y. Nuclei of symbiotic arbuscular mycorrhizal fungi as revealed by in vivo two-photon microscopy. Protoplasma. 1999 Mar;209(1):77-89.

57 Rayner AD. The challenge of the individualistic mycelium. Mycologia. 1991 Jan 1;83(1):48-71.

58 Kokkoris V, Hart M. In vitro propagation of arbuscular mycorrhizal fungi may drive fungal evolution. Frontiers in microbiology. 2019 Oct 22;10:2420.

59 Gehrmann T, Pelkmans JF, Ohm RA, Vos AM, Sonnenberg AS, Baars JJ, Wösten HA, Reinders MJ, Abeel T. Nucleus-specific expression in the multinuclear mushroom-forming fungus Agaricus bisporus reveals different nuclear regulatory programs. Proceedings of the National Academy of Sciences. 2018 Apr 24;115(17):4429-34.

60 Zeng T, Holmer R, Hontelez J, te Lintel-Hekkert B, Marufu L, de Zeeuw T, Wu F, Schijlen E, Bisseling T, Limpens E. Host-and stage-dependent secretome of the arbuscular mycorrhizal fungus Rhizophagus irregularis. The Plant Journal. 2018 May;94(3):411-25.

61 Bécard G, Fortin JA. Early events of vesicular–arbuscular mycorrhiza formation on Ri T-DNA transformed roots. New phytologist. 1988 Feb;108(2):211-8.

62 St-Arnaud M, Hamel C, Vimard B, Caron M, Fortin JA. Enhanced hyphal growth and spore production of the arbuscular mycorrhizal fungus Glomus intraradices in an in vitro system in the absence of host roots. Mycological research. 1996 Mar 1;100(3):328-32.

63 Fauchery L, Uroz S, Buée M, Kohler A. Purification of fungal high molecular weight genomic DNA from environmental samples. InFungal Genomics 2018 (pp. 21-35). Humana Press, New York, NY.

64 Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nature biotechnology. 2019 May;37(5):540-6.

65 Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020 May 1;36(9):2896-8.

66 Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome research. 2017 May 1;27(5):737-46.

67 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015 Oct 1;31(19):3210-2.

68 Smit AF, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015.

69 Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome research. 2009 Sep 1;19(9):1639-45.

70 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature biotechnology. 2019 Aug;37(8):907-15.

71 Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907. 2012 Jul 17.

72 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.

73 Helgason T, Daniell TJ, Husband R, Fitter AH, Young JP. Ploughing up the wood-wide web?. Nature. 1998 Jul;394(6692):431-.

74 Simon L, Lalonde M, Bruns TD. Specific amplification of 18S fungal ribosomal genes from vesicular-arbuscular endomycorrhizal fungi colonizing roots. Applied and environmental microbiology. 1992 Jan;58(1):291-5.

75 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology. 2012 May 1;19(5):455-77.

76 Hoagland DR, Arnon DI. The water-culture method for growing plants without soil. Circular. California agricultural experiment station. 1950;347(2nd edit).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.xlsx
- AdditionalFile10FigureS2.pdf
- AdditionalFile11FigureS3.pdf
- AdditionalFile12FigureS4.pdf
- AdditionalFile13FigureS5.pdf
- AdditionalFile14FigureS6.pdf
- AdditionalFile15TableS1.docx
- AdditionalFile16TableS2.docx
- AdditionalFile2.vcf
- AdditionalFile3.vcf
- AdditionalFile4.txt
- AdditionalFile5.vcf
- AdditionalFile6.vcf
- AdditionalFile7.vcf
- AdditionalFile8.xlsx
- AdditionalFile9FigureS1.pdf