Purifying Selection and Persistent Polymorphism among Nuclei in the Multinucleate Arbuscular Mycorrhizal (AM) Fungi

David Manyara [1]

Marisol Sánchez-García [1, 2]

Markus Hiltunen Thorén [3, 4, 5]

Mercè Montoliu-Nerin [1, 6]

George Cheng [1, 7]

James D. Bever [8]

Hanna Johannesson [4, 5]

*Anna Rosling [1]

1. Department of Ecology and Genetics, Evolutionary Biology, Uppsala University, Uppsala, Sweden.
2. Uppsala Biocentre, Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden.
3. Department of Organismal Biology, Systematic biology, Uppsala University, Uppsala, Sweden.
4. Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden
5. The Royal Swedish Academy of Sciences, Stockholm, Sweden
6. Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden.

7. Department of Molecular Sciences, Microbial biotechnology, Swedish University of Agricultural Sciences, Uppsala, Sweden.

8. Department of Ecology and Evolutionary Biology, and Kansas Biological Survey, University of Kansas, Lawrence, Kansas, USA.

* Corresponding authors: Anna.Rosling@ebc.uu.se;

Abstract

Arbuscular mycorrhizal (AM) fungi form obligate symbiosis with the roots of the majority of land plants and are found in all terrestrial ecosystems. The source and structure of genetic variation in AM fungi has remained an enigma due to difficulties in the axenic cultivation and generation of high-quality genome assemblies from most species. Furthermore, how AM fungi survives long-term without a single nuclear stage per cell life history is puzzling, prompting hypotheses on selection at the nuclear level which functions to purge deleterious mutations. In this study, we aimed to characterize both intra- and inter-organismal genetic variation in AM fungi by analyzing genomic information from individual nuclei of three strains from two species of the genus *Claroideoglomus*. We observed overall low levels of genetic variation within the strains, most of which represent rare variants likely kept at low frequency by purifying selection. We also observed variants that have been maintained as polymorphic across both strains and species. The results in this study affirm our conceptual understanding that nuclei in AM fungal strains function as populations of asexually reproducing units. Further, we propose that selection acts on different levels within the organism, with strong signals of purifying selection on nuclei within strain.

Keywords: AM fungi, Genetic variation, Selection, Single nuclei, Spore, Fungal genomics

Introduction

Most multicellular organisms go through a single-cell bottleneck at some point during the life cycle. This feature is expected to have evolved to reduce genetic conflict, which is likely to occur when genetically diverged cell lineages coexist within a multicellular body and inevitably will reduce fitness of the individual (Richard 1996). At the same time, genetic variation within organisms has been shown to occur in all major branches of multicellular life (Gill et al. 1995; Santelices 1999; Pineda-Krch and Lehtilä 2004). These latter findings challenge the concept of the individual as a single unit of selection (Santelices 1999) and arguments have been raised that within-organism variation benefits the organism, either by providing phenotypic flexibility to a changing environment (Jinks 1952; Pineda-Krch and Lehtilä 2004) and/or by allowing purging of deleterious cell lineages (Otto and Orive 1995).

In this study, we investigate the extent of intra-organismal genetic variation in Arbuscular Mycorrhizal (AM) fungi and infer the processes that have shaped it. AM fungi are obligate plant symbionts that complete their life cycle below ground. They propagate with asexual spores, each harboring hundreds to thousands of nuclei (Bonfante and Genre 2010; Sędzielewska et al. 2011). The mycelium of AM fungi is coenocytic, and bidirectional movement ensures the mixing of the haploid nuclei in the mycelium (Jany and Pawlowska 2010; Sanders and Croll 2010). Unlike most eukaryotes, no single-nuclear stage has been observed in AM fungi, and yet they have persisted and successfully diversified for over 450 million years (Redecker, Kodner, and Graham 2000). Given the lack of a single-nuclear stage and the large number of nuclei that co-exist in mycelia and spores, high genetic variation would be expected among nuclei of AM fungi. Nevertheless, nuclei in AM fungi have been reported to have low genetic divergence, with Single

Nucleotide Polymorphism (SNP) density ranging from about 0.06 to 0.43 SNPs/kb (referred to in these studies as homokaryosis) (Tisserant et al. 2013; Morin et al. 2019). The relatively few variable sites have been suggested to most likely result from mutations (Tisserant et al. 2013). It has been shown that AM fungal strains undergo hyphal fusion and nuclear mixing (Giovannetti, Azzolini, and Citernesi 1999; Bever and Wang 2005; Croll and Sanders 2009), and hence, we can also expect heterokaryotic strains to emerge in which divergent nuclei co-exist. Genomic analyses of the putative mating-type (*mat*) locus in the model AM fungal species *Rhizophagus irregularis* has demonstrated that certain strains carry two alleles for the *mat* locus (Ropars et al. 2016). These are referred to as dikaryotic, and have higher SNP densities ranging from 0.45 to 0.8 SNPs/kb compared to the homokaryotic strains with one *mat* allele (Ropars et al. 2016; Morin et al. 2019). While two types of nuclei in a dikaryon, each carrying one of the *mat* alleles, explains higher observed levels of intra-organismal genetic variation in these particular strains, it remains unresolved how the low genetic variation is maintained in the homokaryotic AM fungal strains that have been studied thus far (Ropars et al. 2016; Morin et al. 2019; Serghi et al. 2021).

By investigating key characteristics of the growth and development in *Claroideoglomus etunicatum* (previously known as *Glomus etunicatum*), Jany and Pawlowska (2010) proposed the hypothesis that the evolutionary longevity of AM fungi results from natural selection on nuclei facilitated by their mode of sporulation. Extensive nuclei mixing and random migration from parts or all of the mycelium into the developing spore provides a unique mode of sporulation that prevents a single founder nucleus and maintains a large nuclear population size across generations (Jany and Pawlowska 2010; Marleau et al. 2011). Together this allows for purifying selection to act against nuclei carrying deleterious mutations. In the study by Jany and

Pawlowska (2010), no nuclear degradation was observed during spore formation and the authors concluded that nuclei do not replicate during spore formation. This trait could however be different in other species as observed by Marleau and colleagues (2011) where nuclear replication during spore formation was observed in *R. irregularis* (previously known as *Glomus irregulare*). Based on the observations above, we build the current study on the conceptual hypothesis that purifying selection acts on nuclei in the developing mycelia and that the nuclear composition in individual spores of AM fungi represents that of the mycelium from which the spore developed.

First, we capitalized on the most recently published full genome assemblies of one strain of *Claroideoglomus claroideum* and two strains of *Claroideoglomus candidum* (Montoliu-Nerin et al. 2020; 2021), and generated single nuclei assemblies for the two *C. candidum* strains to study intra-strain, intra-species, and inter-species genetic variation in this group. We employed phylogenetic network analyses to establish the relatedness among nuclei within and between strains and species, and inferred shared polymorphic sites as ancestral polymorphic sites. Further, we took advantage of the hierarchical sampling to test the hypothesis of selection acting on nuclei within AM fungi by contrasting synonymous and non-synonymous variants at polymorphic sites.

Materials and methods

Fungal strains

Whole inoculum, i.e., soil with both spores and mycelia, of each of *C. claroideum* (strain SA101) and *C. candidum* (strain NC172) was obtained from the International Culture Collection of (Vesicular) Arbuscular Mycorrhizal Fungi (INVAM), and a second strain of *C. candidum* (CCK pot B6-9) was obtained from James Bever's collection at the University of Kansas. These three strains are hereafter referred to as *C. claroideum* SA101, *C. candidum* NC172 and *C. candidum* CCK, respectively. All three strains were isolated from spores collected in the field and subsequently maintained in pot cultures since the early 1990s: The *C. claroideum* SA101 strain was isolated from a wheatfield in Saskatchewan, Canada (Talukdar and Germida 2011), while the two *C. candidum* strains were isolated in consecutive years from an old field (i.e., formerly cultivated but later abandoned) site outside Durham (NC, USA) (Bever et al. 1996; Furrazola et al. 2010).

Spore isolation, nuclei sorting, and genome sequencing

From a single spore of each of the three strains, 24 separate nuclei were sequenced in previous studies. *C. claroideum* SA101 was sequenced by Montoliu-Nerin and colleagues (2020), and used in that study for development of nuclear sorting and assembly workflows. Following their established workflow, *C. candidum* CCK and *C. candidum* NC172 were sequenced by Montoliu-Nerin and colleagues (2021) as a part of a phylogenetic analysis of AM fungi. In brief, for each strain, individual spores were extracted by wet sieving from the whole inoculum. After rinsing the spores, one spore was manually transferred using a glass pipette to an eppendorf tube and crushed in 30 µl ddH2O + 50 µl 1X PBS. DNA in the solution was stained, and particles were

separated by Fluorescence-Activated Cell Sorting (FACS) at the SciLifeLab Microbial Single Cell Genomics Facility with a MoFloTM Astrios EQ sorter (Beckman Coulter, USA). Particles that were identified as nuclei based on fluorescence level and particle granularity were individually sorted into a 96-well plate. Whole-genome amplification was then performed using multiple displacement amplification (MDA). PCR with specific fungal and bacterial rDNA primers was performed to verify that the amplified DNA was of fungal origin, and 24 nuclei from each strain were sequenced by Illumina HiSeq-X at the SNP&SEQ Technology Platform in Uppsala at the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory.

Assembly of reads from single nuclei and screening for the putative mating-type (*mat*) locus

Single nuclei genome assemblies for the strain *C. claroideum* SA101 are described and published (Montoliu-Nerin et al. 2020). Single nuclei assemblies for *C. candidum* CCK and *C. candidum* NC172 were generated in this study following the single nuclei assembly workflow described in Montoliu-Nerin et al. (2020). Specifically, reads from each nucleus were normalized with BBNorm of BBMap v.38.08 (Bushnell et al. 2014) with a target average depth of 100x. The normalized reads from each nucleus were then assembled with SPAdes v.3.12 (Bankevich et al. 2012) using the default settings. A quantitative assessment of each assembly was done with QUAST v.4.5.4 (Gurevich et al. 2013), and the genome completeness was evaluated with BUSCO v.3 with the fungi_odb9 lineage set and *rhizopus_oryzae* species set (Waterhouse et al. 2018).

We used two approaches to assess whether the investigated strains were homo- or heterokaryotic for the putative *mat* locus. First, the presence of the putative *mat* locus and its allelic variants was

assessed with BLASTn v.2.7.1 (Altschul et al. 1990) in the single nuclei genome assemblies using the nucleotide sequences from the putative *mat* locus of *Claroideoglomus* sp. (MH445374 and MH445375) published by Morin and colleagues (2019) as queries. The nucleotide sequence from the top hits in each single nucleus assembly were then extracted and aligned against each other for comparison with MAFFT v.7.407 (Katoh and Standley 2013). Second, as the single nuclei assemblies were not complete, we aimed to avoid false calling of mating-type homokaryons with the first method by screening all amplified nuclei DNA samples for the putative *mat* locus using PCR. Specific primers for the *mat* region were designed with primer-BLAST (Ye et al. 2012) using the sequences obtained from the previous BLAST searches as the PCR template (table S1). The PCR reactions were performed both for the HD1 and HD2 regions of the *mat* locus, and for the whole *mat* region (table S1), and their products were visualized on a 1.5% agarose gel. Successfully amplified products were purified with Agencourt AMPure XP beads (Beckman Coulter, Beverly, MA) following the manufacturer's protocol, and sent to Eurofins Genomics for Sanger sequencing. Raw sequences were manually inspected and assembled using Sequencher v4.9 (Gene Codes, Ann Arbor, MI).

Mapping of reads to their full genome assemblies

Full genome assemblies and annotations for the three strains were generated in previous studies as described in (Montoliu-Nerin et al. 2020; 2021). In summary, pooled reads from the 24 separately sequenced nuclei were normalized and assembled with SPAdes v.3.12.0 (Bankevich et al. 2012). Repeats and transposable elements (TEs) were predicted in the full genome assemblies using RepeatModeler v1.0.8 (Smit and Hubley 2008) and the repeat libraries from the RepeatModeler were subsequently used to mask the genome assemblies using RepeatMasker

v4.0.7 (Smit, Hubley, and Green 2015). Gene prediction in the repeat-masked full genome assemblies was then done with GeneMark-ES v 4.33 (Ter-Hovhannisyan et al. 2008; Montoliu-Nerin et al. 2020; 2021).

For this study, the predicted coding regions (CDS) of the full genome assemblies were further functionally annotated using FunAnnotate v.1.8.9 (Palmer and Stajich 2020). Following the functional annotation, a putative function could be assigned to 49% out of the 15,882 genes in *C. claroideum* SA101, and in 47% out of the 16,088 and 15,761 genes in *C. candidum* CCK and *C. candidum* NC172 respectively. Taking a conservative approach, subsequent analyses of the distribution of polymorphism and selection were performed only on genes that had a function assigned to them, because such genes are expected to be of higher structural quality than genes without any predicted function. We refer to this subset of genes as functionally annotated genes throughout the rest of the text. For each strain, the 24 sets of paired-end reads representing individual nuclei were mapped to their respective full genome assembly (fig. S1, S2A) using the Burrows-Wheeler aligner (BWA) v.0.7.15 (Li and Durbin 2009) with the BWA-MEM algorithm. The resulting Sequence Alignment Map (SAM) files were converted into Binary Alignment Map (BAM) files using SAMtools v.1.5 (Li et al. 2009). Picard v.2.10.3 (http://broadinstitute.github.io/picard/) was used to mark duplicates, sort the BAM files by coordinates, and to replace read groups.

Estimating gene content representation across samples of amplified single nuclei

To assess whether there is a structure of different nuclei in the same strain carrying different sets of functionally annotated genes, we performed a principal component analysis (PCA) based on

the presence-absence of all functionally annotated genes in each nucleus within each of the three strains. For each strain, read coverages of the functionally annotated genes across all the nuclei were first obtained using bedtools v.2.29.2 (Quinlan 2014). Mapping files for each individual nucleus were used to compute the proportions of the functionally annotated genes covered by reads (i.e, breadth of coverage). A cutoff of 0.1 was then implemented to generate a presence-absence matrix, with proportions of less than 0.1 scored as 0 for absence and proportions equal to or greater than 0.1 scored as 1 for presence. PCAs and histograms to visualize patterns and distribution of the functionally annotated genes in each nucleus were generated from the presence-absence matrix in R v.4.1.1 (R Core Team 2021).

High quality chromosome-level genome assemblies of the model AM fungus *R. irregularis* recently demonstrated that its genome is organized into two categorized compartments: A and B, with the A-compartment containing most core-genes and the B-compartment carrying dispensable genes and more repeats (Yildirir et al. 2022). To assess if homologous genes of both of these compartments were represented in the full genome assemblies of the three strains in our study, we assessed the presence of genes assigned to the two compartments. A MegaBlast search was performed on the nucleotide sequences of all the predicted genes in the three strains using the nucleotide sequences of all *R.irregularis* DAOM197198 compartment A and B genes as queries (https://zenodo.org/record/5181509#.Yxhz9uxBzUI; G. Yildirir personal communication). The analysis was done using BLAST v.2.12.0 (Altschul et al. 1990) and only top hits with an e-value of 1E-5 or lower were considered a match. The percentage of genes potentially representing the two compartments A and B was obtained from the division of the total number of hits to the number of all genes within a strain.

SNP calling and analyses of genetic variation within strains

Following read alignment, the Genome Analysis Toolkit (GATK) v.4.1.7 (McKenna et al. 2010) was used to call SNPs between the mapped reads and the full genome assemblies. For each strain, the HaplotypeCaller of the GATK pipeline was first run in the HaplotypeCaller Reference Confidence Model (GVCF) mode on each nucleus with the following parameters; ploidy 1, minimum-mapping-quality 30, mbq 20, and max-alternate-alleles 1 (maximum number of alternate alleles to haplotype). Genotyping was then performed individually on all the resultant genomic variant call format (GVCF) files with an additional parameter - stand-call-conf 10 - to retain only variant sites with a minimum Phred-scaled confidence threshold of ten. All the non-variant sites were subsequently excluded and only SNP variants were selected and hard-filtered with the following parameters; QD <2 (SNP confidence normalized by depth), FS >60 (to minimize strand bias at the SNP site), MQ <40 (root mean square mapping quality of a SNP), MQRankSum <-12.5 (Rank Sum Test for mapping qualities of the reads supporting the full allele and the alternate allele), ReadPosRankSum <-8 (u-based z-approximation from the Rank Sum Test for site position within reads), QUAL <30 (filter out SNPs with a quality score of less than 30), and SOR > 3.0 (remove the long tail of SNPs that show sequencing bias where one strand is favored over the other). To retain high quality SNPs and to minimize noise potentially caused by MDA, sequencing errors, and mapping errors for our downstream analyses, we implemented an additional filter that removed intra-nucleus polymorphic sites by requiring all genotypes to be supported by at least five reads, and have an allele fraction equal to or above 0.9 for the alternate allele and equal to or less than 0.1 for the reference allele within each nucleus. The remaining genotypes were scored as missing data. For each strain, one SNP dataset was then generated across the 24 nuclei by including all sites that were supported with data from at least 16 of the 24

nuclei (i.e., 67%). Alternate allele frequencies were computed using the vcfR v.1.14.0 R package (Knaus and Grünwald 2017) by considering the total of the alternate alleles divided by the total of both alternate and reference alleles across all the nuclei within a strain and plotted in R v.4.1.1 (R Core Team 2021). Based on the gene predictions, the whole-genome SNP within-strain datasets were further partitioned into the non-repetitive fraction, the coding fraction (CDS), and the functionally annotated coding fraction. SNP density was calculated for non-repetitive and the functionally annotated CDS fractions separately as the total number of SNPs divided by the total number of nucleotides in the respective genomic fraction (fig. S2A).

Analysis of fixed differences and shared polymorphism across species and strains

To analyze fixed differences and the distribution of polymorphism between species and strains, we first performed a joint SNP calling of all 72 nuclei of the three strains by mapping all reads to a common reference genome *C. claroideum* SA101 using the same settings as described in the SNP calling section above. For all downstream analyses, only SNPs supported with data from at least 48 of the total 72 nuclei (i.e., 67%) in the functionally annotated CDS were considered. SNPs located within TEs were also removed to minimize potential bias in the results. Such SNPs were identified by intersecting the SNP set and the annotation file for the *C. claroideum* SA101 strain using the intersect function of bedtools v.2.29.2 (Quinlan 2014), then removing any site annotated as “retrotransposons” and “transposable”. The SNP dataset was then split into strain-specific datasets using BCFtools v.1.12 (Danecek and McCarthy 2017). Each of the three strain-specific dataset was filtered to extract polymorphic sites using a custom script. Finally, a custom script was used to generate shared SNPs, i.e., sites that were polymorphic in all three strains or

274 polymorphic in two strains and fixed in the third strain, and strain specific sites that were
275 polymorphic in one strain and fixed in the other two strains (fig. S2B).
276
277 To obtain a larger inter-species dataset for the shared and the species-specific polymorphism, we
278 performed another mapping with only *C. claroideum* SA101 and *C. candidum* CCK strains, as
279 these two were represented by better datasets than the *C. candidum* NC172 strain. In this
280 analysis, all reads from *C. claroideum* SA101 and *C. candidum* CCK strains were mapped to the
281 *C. claroideum* SA101 full genome assembly and joint SNP calling and filtering was performed
282 as described above, excluding SNPs located within TEs, and including SNPs only if they were
283 supported by data from at least 32 of the total 48 nuclei (i.e., 67%). Using only the SNPs in the
284 functionally annotated CDS, we analyzed the shared polymorphism in the two species as well as
285 the fixed differences between the two species. BCFtools v.1.12 (Danecek and McCarthy 2017)
286 was used to split the inter-species SNP dataset into strain-specific datasets and custom scripts
287 were used to separately extract shared polymorphic SNPs and fixed SNPs. The fixed differences
288 between the two species were computed as sites where alternate alleles were fixed in the *C.*
289 *candidum* CCK strain but reciprocally fixed for the reference allele in the *C. claroideum* SA101
290 strain (fig. S2C).
291
292 Finally, we generated an intra-species dataset to expand the analysis of intraspecific genetic
293 variation: to minimize the likelihood that decreased mapping frequency to the full genome
294 assembly of the sister species would mask shared polymorphism and fixed differences between
295 the two *C. candidum* strains, we mapped all reads from *C. candidum* CCK and *C. candidum*
296 NC172 strains to the *C. candidum* CCK full genome assembly. Joint SNP calling and filtering

was then performed as described above, including SNPs only if they were supported with data from at least 32 of the 48 nuclei (i.e., 67%) and excluding SNPs located within TEs. Using only the SNPs in the functionally annotated CDS, fixed differences and shared polymorphism between the *C. candidum* strains were evaluated. Fixed differences between the *C. candidum* strains were calculated as the number of sites that were fixed alternate alleles in the *C. candidum* NC172 strain but reciprocally fixed for reference alleles in the *C. candidum* CCK strain (fig. S2D).

Alternate allele frequencies were computed separately for the different datasets and different genome fractions by considering the total number of alternate alleles divided by the total number of both alternate and reference alleles across all nuclei within a strain. This was done using vcfR v.1.14.0 R package (Knaus and Grünwald 2017) and plotted in R v.4.1.1 (R Core Team 2021).

Phylogenetic relationship and genetic structure in strains and species

Phylogenetic networks were used to infer the relationship between nuclei within strains based on SNPs in the non-repetitive fraction of the genome, as identified when reads were mapped to their own full genome assemblies. For this purpose, Variant Call Format (VCF) files were first converted to the nexus format using the vcf2phylip v.2.0 script (Ortiz 2019) before neighbor-net construction in SplitsTree v.4.14.8 (Huson and Bryant 2006). To visualize the relationship between nuclei across the three strains, a phylogenetic network was also generated from the 2,323 SNPs within the functionally annotated CDS when all three strains were mapped to the *C. claroideum* SA101 full genome assembly.

The fixation index (F_{ST}) summary statistic (Weir and Cockerham 1984), calculated using vcftools v.0.1.16 (Danecek et al. 2011), was used to estimate the relationship and genetic differentiation between pairs of strains and between species. Thus, the F_{ST} statistic was computed separately for the *C. candidum* strains when they were both mapped to the *C. candidum* CCK full genome assembly, and for the species comparison between *C. claroideum* SA101 and *C. candidum* CCK when they were both mapped to the *C. claroideum* SA101 full genome assembly.

Signatures of selection on nuclei

To assess signatures of selection shaping diversification in *Claroideoglomus*, we computed the ratio of the number of non-synonymous variants per non-synonymous sites to the number of synonymous variants per synonymous sites (dN/dS) using a custom script. Generally, dN/dS ratios that are less than 1 indicate purifying selection and the relative strength of selection within and across strains can be assessed using this metric, with lower ratios implying higher levels of purifying selection (Kryazhimskiy and Plotkin 2008). dN/dS ratios were computed for each nucleus separately using SNPs within the functionally annotated CDS when reads were mapped to their own full genome assemblies (intra-strain mapping), as well as the shared SNPs and fixed differences between strains and species identified from the intra-species and inter-species mapping respectively. One-way analysis of variance (ANOVA) followed by Tukey's HSD statistical tests in R v.4.1.1 (R Core Team 2021) were used to test for significant differences in levels of selection.

Evaluation of strength of selection

To further examine the strength of selection, we computed the proportions of SNP variants with different potential functional impacts on encoded proteins as predicted by the software package SnpEff v.5.0e (Cingolani et al. 2012). SnpEff annotates each SNP variant and infers its effects on known encoded proteins as either high, moderate, low, or no impact. SNPs with high impact correspond to non-synonymous variants that disrupt protein function, moderate impact SNPs are non-synonymous variants that may change protein sequences without disrupting their function, and low impact SNPs are synonymous (silent) variants with a low effect on proteins and are mostly harmless. Using only the functionally annotated CDS, SNPs shared between species and strains were scored by their different impact categories. The annotations for the *C. claroideum* SA101 strain and the *C. candidum* CCK strain were used as custom annotation databases for the respective analysis. We also calculated the corresponding expected proportions based on the number of possible changes for the three categories using a custom script. Alternate allele frequencies for the three different categories were finally generated and plotted in R v.4.1.1 (R Core Team 2021).

Results

Assembly quality and gene content representation within nuclei and strains

Single-nucleus genome assemblies were generated from the randomly amplified and sequenced DNA. Across the three strains, individual nuclei assemblies varied in size, from 2.2 - 52.6 Mbp (table S2-S4), while the full genome assemblies for the three strains ranged from 68.1 - 69.9 Mbp (table S5). The completeness of nuclei assemblies, as estimated by the percentage of fragmented and complete BUSCO genes, also varied. Across the single nuclei assemblies, the

364 strain *C. claroideum* SA101 showed an average completeness of 49%, ranging from 20-80%
365 (table S2). For the two *C. candidum* strains, the corresponding number was 57% (ranging from
366 4-78%) for *C. candidum* CCK while an overall lower completeness was recorded for *C.*
367 *candidum* NC172 with an average at 27% (ranging from 9-56%) (table S3-S4). For the full
368 genome assemblies of the three strains, the completeness was 91-92 % (table 5). Mapping reads
369 of single nuclei to their corresponding full genome assemblies demonstrated that nearly all reads
370 from each nucleus were represented in the full assembly with an average read mapping of 99.5%
371 for *C. claroideum* SA101, 99.6% for *C. candidum* CCK, and 99.3% for *C. candidum* NC172
372 (table S2-S4). Further, assembled reads from single nucleus samples covered on average 49-76%
373 of their corresponding strain full assemblies (table S2-S4).
374
375 Gene content across single-nucleus assemblies was explored by scoring the presence-absence of
376 the 7,400-7,700 functionally annotated genes in each strain. Between 39% and 45% of all
377 functionally annotated genes were present in each of the 24 single nuclei assemblies in *C.*
378 *claroideum* SA101 and *C. candidum* CCK strains, respectively. In the *C. candidum* NC172
379 strain, however, only 2.5% were present in each nuclei (fig. S3). Furthermore, the majority of the
380 functionally annotated genes were represented by reads from most single nuclei within
381 the strains: 85% and 89% were present in 20 or more nuclei in *C. claroideum* SA101 and *C.*
382 *candidum* CCK strains respectively, and the corresponding number was 16% in *C. candidum*
383 NC172 (fig. S3). These results, together with the single nuclei assembly statistics (table S2-S4),
384 indicate that data quality was generally lower for the *C. candidum* NC172 strain compared to the
385 *C. claroideum* SA101 and *C. candidum* CCK strains. For all three strains, a PCA of
386 presence/absence of the functionally annotated genes showed no structure among nuclei with

387 regards to gene content (fig. S4). For *C. claroideum* SA101 and *C. candidum* CCK, data from
388 single nuclei clustered tightly with the exception of a couple of nuclei which were disparate in
389 the plot, consistent with observed small single nuclei assembly size and low full-assembly
390 coverage for these nuclei (fig. S4A-S4B; table S2-S3). For the *C. candidum* NC172 strain, on the
391 other hand, data from separate nuclei were more scattered in ordination space, consistent with the
392 overall smaller single nuclei assemblies and lower coverage for this strain (fig. S4C; table S4).
393
394 We also assessed whether genes from both compartments A and B, as identified in *R. irregularis*
395 DAOM 197198 (Yildirir et al. 2022), were represented in the full assemblies of the three
396 *Claroideoglomus* strains. Across the three full assemblies, the majority of genes could be
397 assigned to either of the putative compartments A (46-47% of the total number of genes) or B
398 (15-16%), while 37-39% of genes remained unassigned (table S6). While we do not know if
399 there is compartmentalization, and if so, the actual proportion of A and B compartments in
400 genomes of *Claroideoglomus* species, our observation confirms that genes matching those of
401 *Rhizophagus* in both compartments are likely represented in the data. Together, the reference
402 assembly statistics, read coverage and gene content confirms that the full genome assemblies for
403 the three strains are fairly complete and that single nuclei read data is suitable for scoring
404 variance across nuclei.
405
406 **Nuclei are haploid and strains are monokaryotic for the putative mat locus**
407 The distribution of intra-nucleus allele fractions was calculated after mapping single-nucleus
408 reads to their corresponding full genome assembly. All nuclei in the three strains had an allele

fraction distribution that formed a distinct fraction peak at 1, strongly supporting the assumption that nuclei are haploid (fig. S5-S7). The observation of low frequency alleles within nuclei was assumed to represent errors caused by MDA, sequencing or mapping. The putative *mat* locus was identified in 20 of the 24 nuclei assemblies for both *C. claroideum* SA101 and *C. candidum* CCK strains, and in 11 out of 24 nuclei in the *C. candidum* NC172 strain (table S2-S4). All the retrieved putative *mat* allele sequences were identical across nuclei in each strain and between the two *C. candidum* CCK strains. Together, a consistent allele fraction peak at 1 in all individual nuclei and the presence of identical sequences at the putative *mat* locus across nuclei in each strain demonstrates that the three strains carry haploid nuclei and that they are monokaryotic for the *mat* locus.

Low genetic variation among nuclei is dominated by rare variants

To assess the level of polymorphism across nuclei within each of the three strains, we used filtered SNP datasets and scored each nucleus as either reference or alternate alleles for all sites that were supported with enough reads. Within the non-repetitive fraction, a total of 16,110 SNPs were identified in *C. claroideum* SA101, 18,790 SNPs in *C. candidum* CCK, and 9,811 SNPs in *C. candidum* NC172 (table S5). These numbers translated to a SNP density of 0.35 SNP/kb, 0.40 SNP/kb, and 0.21 SNP/kb respectively. Within the functionally annotated CDS, the corresponding SNP density was 0.18 SNP/kb, 0.28 SNP/kb, and 0.17 SNP/kb for *C. claroideum* SA101, *C. candidum* CCK, and *C. candidum* NC172 respectively (table S5). Further, the alternate allele frequency distribution for SNPs in the whole genome had a right-skewed distribution with a dominant peak close to zero, in all strains (fig. S8), indicating that the majority of the SNPs were rare variants. Similar alternate allele frequency distribution patterns

were consistently observed and recovered for polymorphic sites in both the non-repetitive and the CDS fractions of the genome (fig. S8). For all strains, phylogenetic network analyses based on SNPs detected in the non-repetitive fractions of the full genome assemblies revealed little structure with individual nuclei separating on long branches in the networks (fig. S9-S11). The observed lack of structure in the networks, coupled with the low degrees of reticulation, indicated that the majority of the observed polymorphism within the strains represents variants unique to each nucleus. However, some reticulation was observed between the nuclei in the networks, indicating that some SNPs are shared among nuclei within the strains and suggest different topologies (fig. S9-S11).

Fixed differences and shared polymorphism between species and strains

To explore fixed differences and the distribution of polymorphism between species and strains, we evaluated (1) the fixed differences between species and strains; and (2) the distribution of polymorphic sites across species and strains. For the fixed differences analysis, the between-species analysis based on the inter-species dataset revealed a total of 630 SNPs that were fixed differences between the *C. claroideum* and the *C. candidum* species (fig. S2C). On the other hand, only two fixed SNP sites were identified between the two *C. candidum* strains in the intra-species mapping (fig. S2D) thus indicating a very low difference between *C. candidum* CCK and *C. candidum* NC172 strains.

Our analyses also revealed that there was a substantial number of shared polymorphism, i.e. sites that were polymorphic in all three *Claroideoglomus* strains, in two species, as well as between the two *C. candidum strains* (fig. S2B-S2D). The first mapping including three strains revealed

455 that 114 SNPs were polymorphic across all three *Claroideoglomus* strains, corresponding to
456 about 5% of the total 2,323 SNPs identified across the three strains (fig. S2B). The alternate
457 allele frequency spectrum of the 114 SNPs shared across all the three strains was shifted slightly
458 to the right with peaks at around 0.2 and 0.3 (fig. 1A) indicating that these variants occurred in
459 intermediate frequencies. In contrast, the strain-specific SNPs (fig. 1C-1E) had a left-skewed
460 allele frequency spectrum consistent with an excess of rare alleles. The 114 SNPs shared across
461 the three strains were detected in 31 genes with functional annotations including roles linked to
462 DNA and RNA binding, as well as regulation of transcription (supplement B). By excluding the
463 *C. candidum* NC172 strain which had an overall lower data quality, the inter-species analysis
464 revealed 363 SNPs that were polymorphic across the two species (fig. S12). This corresponded
465 to about 9% of the total 4,111 SNPs identified across the two species. Again, the SNPs shared by
466 the two species were detected at higher frequencies compared to the species-specific SNPs in this
467 dataset (fig. S12). The 363 SNPs shared between the two species were detected in 83 genes
468 (supplement B) with functional annotation similar to those of the smaller gene set including three
469 strains discussed above. In the intra-species analysis, a total of 959 SNPs were polymorphic
470 across the two *C. candidum* strains and this corresponded to about 33% of the total 2,947 SNPs
471 identified across two strains. The shared SNPs also appeared at slightly higher frequencies
472 compared to the strain-specific SNPs, the majority of which appeared at lower frequencies (fig.
473 S13).

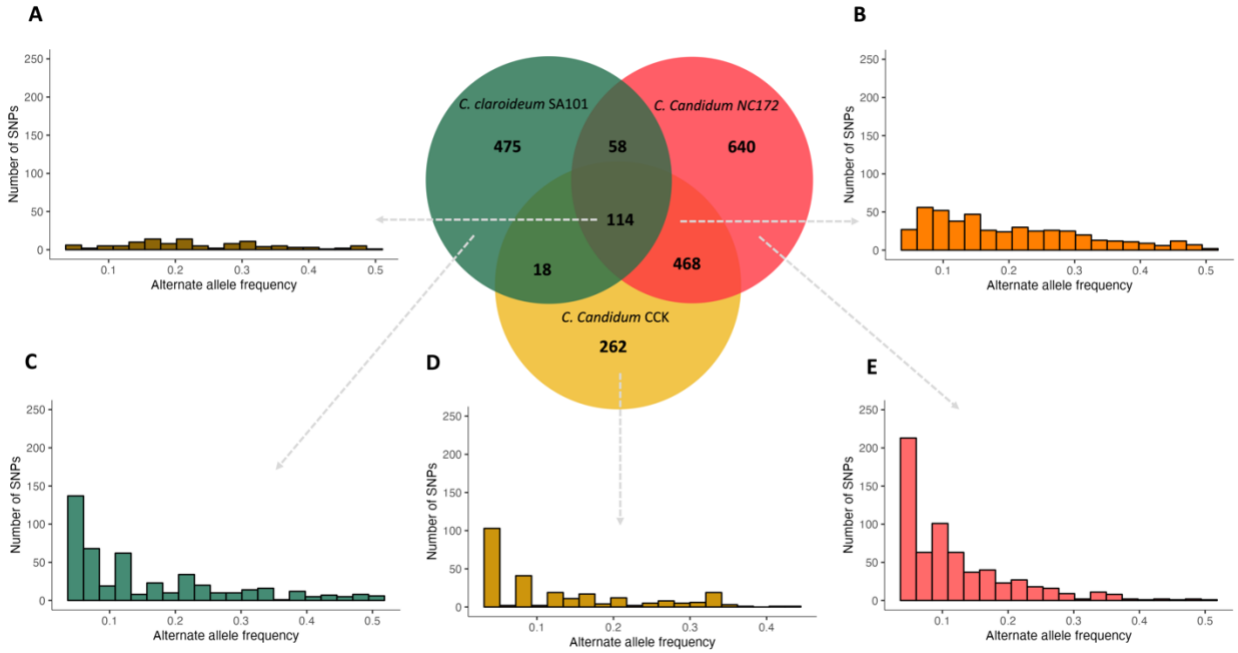
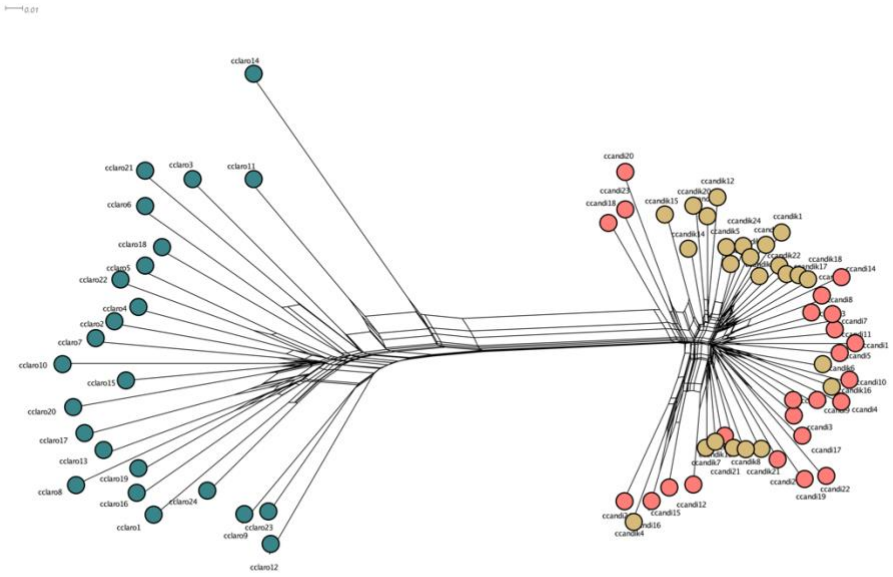


Figure 1. Venn diagram illustrating the overlap of SNPs in the three strains and their corresponding alternate allele frequencies. A, Histogram of alternate allele frequency distribution for the shared SNPs between the three strains. B, Histogram of alternate allele frequency distribution for the shared SNPs between the two *C. candidum* strains. C, Histogram of alternate allele frequency distribution for the SNPs specific to *C. claroideum* SA101. D, Histogram of alternate allele frequency distribution for the SNPs specific to *C. candidum* CCK. E, Histogram of alternate allele frequency distribution for the SNPs specific to *C. candidum* NC172.

Nuclei within the two C. candidum strains are closely related, and phylogenetically distinct from those in C. claroideum

A phylogenetic network based on the 2,323 SNPs identified in the first mapping that included the three strains visualized how nuclei were related across strains (fig. 2). As expected, nuclei from the two *C. candidum* strains clustered together and were mixed in the network while distinctly separating from nuclei in the *C. claroideum* SA101 strain. Further, estimation of genetic differentiation in a pairwise comparison revealed low genetic differentiation between two *C.*

490 *candidum* strains as indicated by an F_{ST} value of 0.13 while high levels of genetic differentiation
491 were observed between C. *claroideum* SA101 and *C. candidum* CCK strains, with an F_{ST} value
492 of 0.68. This was expected as the two *C. candidum* strains differed by only a couple of fixed
493 SNPs compared to the many fixed differences between *C. claroideum* SA101 and *C. candidum*
494 CCK. The observed reticulation separating the two species in the phylogenetic networks could be
495 a result of ancestral polymorphisms or hybridization between species, and shared polymorphisms
496 are maintained across species after diversification.



497

498 **Figure 2**. Phylogenetic network of the 2,323 shared SNPs in the functionally-annotated coding fractions
499 of the three Claroideoglomus strains - *C. claroideum* SA101 (denoted as cclaro, highlighted by
500 aquamarine-colored circles), *C. candidum* CCK (denoted as ccandik, highlighted by dark goldenrod-
501 colored circles) and *C. candidum* NC172 (denoted as ccandi, highlighted Indian-red-colored circles) when
502 mapped to the *C. claroideum* SA101 full genome assembly. Each strain is represented by reads from 24
503 single nuclei (numbered 1-24).

504

Signatures of purifying selection within and between strains and species

To understand how genetic variation is maintained in the strains, we assessed signatures of selection within and between the strains and species. The overall selection pressure on individual nuclei was inferred by the dN/dS ratio for SNPs in the functionally annotated CDS. We first estimated selection within the strains based on SNPs when reads were mapped to their respective full genome assemblies. In all three strains, an overall signature of purifying selection was detected as indicated by a dN/dS ratio lower than 1. Levels of dN/dS were significantly different within the three strains (Tukey's HSD, p-value < 0.001; fig. 3), with dN/dS ratios of 0.34 in *C. claroideum* SA101, 0.40 in *C. candidum* CCK, and 0.44 in *C. candidum* NC172 (fig. 3).

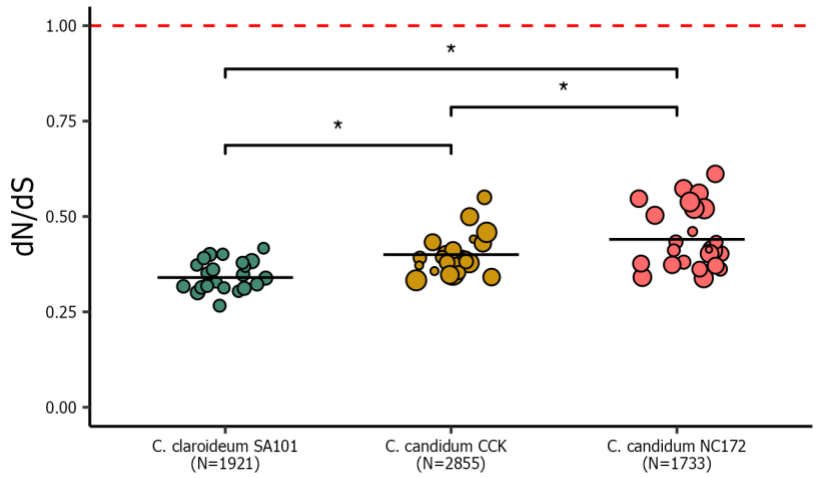


Figure 3. Comparison of signatures of selection within the *Claroideoglomus* strains. dN/dS ratios were computed for SNPs in the functionally-annotated coding fractions of the three strains - *C. claroideum* SA101 (left panel), *C. candidum* CCK (middle panel), and *C. candidum* NC172 (right panel) when mapped to their respective full genome assemblies. Each point represents the dN/dS ratio for all genes that carried a polymorphism in a single nucleus and each strain has 24 single nuclei. Point sizes correspond to the counts of SNPs in nuclei and the number of SNPs in each panel is indicated as N. The red line represents the expected value on the basis of selective neutrality.

523 We also detected significantly different levels of dN/dS when we estimated selection between
524 species and strains. From the inter-species analysis represented by *C. claroideum* SA101 and *C.*
525 *candidum* CCK strains, we observed a significantly higher signature of purifying selection in the
526 species-specific polymorphism with an average dN/dS ratio of 0.33, compared to the
527 polymorphism shared across the two species whose average dN/dS ratio was 0.4 (ANOVA, p-
528 value < 0.05; fig 4A). From the *C. candidum* intra-species analysis, we found similar levels of
529 purifying selection in the strain-specific polymorphism and in the polymorphism shared across
530 the two strains with an average dN/dS ratio of 0.47 and 0.49, respectively (ANOVA, p-value >
531 0.05; fig 4B).
532

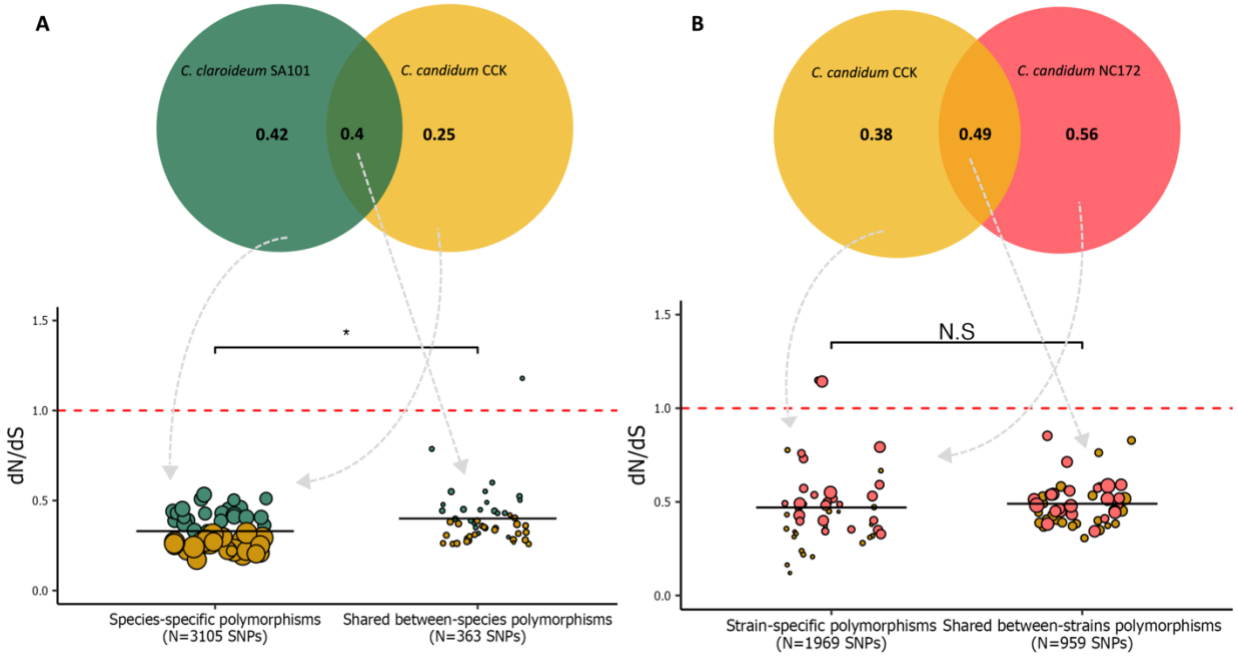


Figure 4. Comparison of signatures of selection within and between species and strains. A, Comparison of signatures of selection within and between the two *Claroideoglomus* species represented by the *C. claroideum* SA101 (aquamarine) and *C. candidum* CCK (dark goldenrod) strains respectively. dN/dS ratios were computed for SNPs in the functionally-annotated CDS of the two strains when mapped to the *C. claroideum* SA101 full genome assembly. The left panel of the scatterplot represents polymorphisms unique to the two species and the right panel represents the polymorphisms shared between the two species. B, Comparison of signatures of selection within and between the two *C. candidum* strains represented by the *C. candidum* CCK (dark goldenrod) and *C. candidum* NC172 (Indian-red) strains respectively. dN/dS ratios were computed for SNPs in the functionally-annotated CDS of the two strains when mapped to the *C. candidum* CCK full genome assembly. The left panel of the scatterplot represents polymorphisms unique to the two strains and the right panel represents the polymorphisms shared between the two strains. In all the scatterplots, each point represents the dN/dS ratio for all genes that carried a polymorphism in a single nucleus and each strain has 24 single nuclei. Point sizes correspond to the counts of SNPs in nuclei and the number of SNPs in each panel is indicated as N. The red line represents the expected value on the basis of selective neutrality.

Finally, we found a strong signature of purifying selection in the 630 fixed inter-species differences with an average dN/dS ratio of 0.32. Taken together, these results indicate that purifying selection operates both within and between species and strains in our study system.

Predicted functional effects of variants impacts the frequency of shared polymorphisms

The annotation of SNPs using SnpEff (Cingolani et al. 2012) allowed us to analyze the variance underlying the detected signatures of purifying selection within species and strains based on the different predicted functional impacts of SNPs on the encoded proteins. Overall, highly deleterious SNPs were found in very low proportions relative to the total number of SNPs. For the polymorphism shared between species, about 3% of the total 363 shared SNPs were highly deleterious, 57.9% were mildly deleterious, and 39.1% were low-impact SNPs (table S7). These observed percentages for the highly deleterious and the mildly deleterious SNPs were significantly lower compared to their expected proportions without purifying selection at 7.5% and 72.2% respectively (Fisher's exact test, p-value < 0.05). The observed proportion of the low-impact SNPs on the other hand was higher than its expected proportion without purifying selection at 20.4%. Similarly, polymorphic sites shared between the two *C. candidum* strains showed the same trend with only about 2% of 969 SNPs being highly deleterious, 49.6% being mildly deleterious, and 48.4% being low-impact (table S7). Again, the observed percentages for the highly deleterious and the mildly deleterious SNPs were significantly lower than their expected proportions without purifying selection at 6.5% and 73.7% respectively (Fisher's exact test, p-value < 0.001). The observed proportion of the low-impact SNPs was higher than its expected proportion without purifying selection at 19.9%. The alternate allele frequency distribution of the SNPs in the different categories of impacts within strains and species reflected

these observed results well with the highly deleterious SNPs occurring in low frequencies (fig. S14), consistent with purifying selection acting more efficiently against highly deleterious mutations. Collectively, these results provide evidence that purifying selection acts to modulate the accumulation of highly deleterious mutations in AM fungal strains. However, it is interesting to note that variants that are predicted to be highly deleterious for gene function are present, although at a very low frequency, within strains and even shared among strains and species. The most abundant gene ontology (GO) annotations for the genes carrying these highly deleterious variants included nucleic acid binding, protein binding, and double-stranded DNA binding molecular functions (supplement B).

Discussion

To date, analysis of genetic diversity across AM fungi has been largely limited to strains of the model genera *Rhizophagus* and *Gigaspora* that grow well axenically (Ropars et al. 2016; Morin et al. 2019; Chen, Morin, et al. 2018; Lin et al. 2014). In this study, we extended beyond the few well-studied taxa by exploiting genome data generated directly from 24 individual nuclei in each of the three strains within the non-model genus *Claroideoglomus*. Our unique sampling strategy enabled us to analyze the partitioning of variation among nuclei, strains and species. In line with previous work (Montoliu-Nerin et al. 2020), we demonstrate that combining data from multiple individually sequenced nuclei results in fairly complete full genome assemblies for all three strains (table S5). Subsequently, mapping reads from individual nuclei allows us to make adequate analyses of genetic variation within and between strains. To that end, by requiring at least five reads for all genotypes, an allele fraction of at least 0.9 to call a SNP, and excluding sites with more than 33% missing data, we implemented a conservative approach to identify and

analyze only high-confidence polymorphisms across nuclei. We investigated how the three *Claroideoglomus* strains were related based on genetic diversity within and between strains. Our findings and their implications for evolutionary mechanisms shaping genetic diversity in AM fungi are discussed below.

Delineation of strains and species of *Claroideoglomus*

Partitioning the SNPs into polymorphic and fixed SNPs allowed us to investigate, at an unprecedented level, the genetic similarities and differences between nuclei within and between strains and species of a group of AM fungi. Our data shows support for the recent phylogenetic analysis indicating that the two *C. candidum* strains are very closely related, with *C. claroideum* clearly separated from the two (Montoliu-Nerin et al. 2021). Specifically, we demonstrate that the two *C. candidum* strains are highly similar with very few fixed differences and many polymorphic sites shared between them. This similarity is not surprising as the two strains were isolated in consecutive years from the same field site and it is likely that the two strains even originate from the same clonal lineage (Bever et al. 1996; Furrazola et al. 2010; Talukdar and Germida 2011). Furthermore, despite the known difficulty in species delineation within *Claroideoglomus* based on rDNA markers (Vankuren et al. 2013; House et al. 2016), the considerable number of fixed differences between *C. claroideum* and *C. candidum* support these being two different species. Phylogenetic network analysis provided additional evidence for these observations with a close clustering of the two *C. candidum* strains and a distinct separation of the two species (fig. 2), and the genetic divergence estimated by the population differentiation parameter F_{ST} revealed a low genetic differentiation between the two *C. candidum* strains and a high genetic differentiation between the two species.

Haploid nuclei evolve independently within strains through the accumulation of mutations

We used within-nuclei alternate allele fraction distribution patterns to firmly demonstrate that all nuclei are haploid in the three analyzed *Claroideoglomus* strains (fig. S5-S7). The use of allele fractions of individual nuclei is a useful method to infer ploidy and our results from this analysis are in line with observations made in earlier studies of ploidy in AM fungi based on single copy genes (Hijri and Sanders 2005) and genome size estimates in dikaryon and monokaryon-like strains of *R. irregularis* (Ropars et al. 2016). Furthermore, only one allele of the putative *mat* locus was identified across nuclei within each of the three strains, suggesting that the strains investigated herein are monokaryotic for mating-type and represent a "non-mated" stage in the proposed AM fungal life cycle (Ropars et al. 2016; Chen, Morin, et al. 2018). SNP density estimates across nuclei within the three strains ranged from 0.17 SNP/kb to 0.28 SNP/kb in the functionally annotated coding fractions of the genomes. This result, and an observed intra-strain allele frequency pattern with a peak close to zero (fig. S8), is in agreement with previous reports of low SNP density in other AM fungal strains that are monokaryotic for the *mat* locus (Ropars et al. 2016; Morin et al. 2019; Lin et al. 2014; Tisserant et al. 2013).

The origin of intra-organismal genetic variation in AM fungi has been the subject of extensive debate. Previous studies have proposed mixing of different nuclei, i.e. nucleotypes, as a result of hyphal fusion (Giovannetti, Azzolini, and Citernesi 1999; Bever and Wang 2005), potentially followed by recombination events identified using polymorphic markers (Vandenkoornhuyse, Leyval, and Bonnin 2001; Gandolfi et al. 2003) and sequencing data (Chen, Mathieu, et al. 2018; Mateus et al. 2022), although these predictions are sensitive to filtering and whether repeat regions are included or not (Auxier and Bazzicalupo 2019). In our study, we detected no

642 phylogenetic structure with nuclei separating on long branches thus indicating that nuclei evolve
643 independently within strains through the accumulation of mutations (fig. S9-S11). We observed
644 minor reticulation between nuclei in the phylogenetic networks but we cannot distinguish if this
645 is a result of recombination or shared ancestral polymorphism. Overall, the confirmation of
646 haploid nuclei that carry mostly unique variants, while all having the same putative *mat* locus
647 allele, suggests that nuclei in these three *Claroideoglomus* strains function as populations of
648 independently replicating asexual units (fig. S9-S11).
649
650 **Purifying selection purges deleterious mutations among nuclei in AM fungi**
651 AM fungi harbor large nuclear populations in their coenocytic mycelia. Observations of
652 asynchronized nuclear division, nuclear degradation in hyphae and extensive nuclear mixing in
653 growing mycelia and during spore formation led Jany & Pawlowska (2010) to propose that the
654 AM fungal mode of sporogenesis represent a life history adaptation that allows for moderating
655 the accumulation of deleterious mutations thereby ensuring the longevity of AM fungi. Our work
656 provides the first quantitative evidence for nuclear-level selection in AM fungi, with strong
657 signatures of purifying selection, as indicated by low dN/dS ratios ranging from 0.34 to 0.44 (fig.
658 3) for polymorphism across nuclei within the three *Claroideoglomus* strains. These results
659 demonstrate that selection acts on individual nuclei to remove deleterious mutations in AM
660 fungi, a process that is consistent with the generally low levels of within-species genetic
661 variation in other AM fungal strains (Lin et al. 2014; Morin et al. 2019; Ropars et al. 2016;
662 Tisserant et al. 2013).
663

664 **Shared polymorphism between strains and species of *Claroideoglomus***

665 A noteworthy finding of our study was the identification of polymorphisms being shared
666 between strains of *C. candidum* and also between *C. candidum* and *C. claroideum*. Previous
667 studies have suggested that genetic polymorphism is maintained in AM fungal strains over
668 experimental-scale growth and cell divisions (Boon et al. 2015; Wyss et al. 2016) and the results
669 presented here support the maintenance of genetic polymorphism at a longer evolutionary time
670 scale. Signatures of purifying selection remain high in polymorphisms shared between species
671 and strains (fig. 4). Notably, we detect higher average dN/dS ratio of the polymorphism shared
672 between species compared to that of the species-specific polymorphisms, indicating that shared
673 polymorphisms are under weaker purifying selection (fig. 4A). Different models, not mutually
674 exclusive, explaining the maintenance of polymorphisms are outlined below.
675
676 One model to explain inheritance and maintenance of polymorphisms over evolutionary time is
677 that incomplete lineage sorting of variation occurs in the large nuclear populations present within
678 each strain during diversification and speciation. It is well established that spores of AM fungal
679 strains carry high numbers of nuclei (Bécard and Pfeffer 1993; Jany and Pawlowska 2010;
680 Marleau et al. 2011; Kokkoris et al. 2020) and the mode of sporogenesis ensures transmission of
681 multiple randomly combined nuclei from the organism (Jany and Pawlowska 2010). Maintained
682 intra-organismal polymorphism during sporogenesis has been observed in *Claroideoglomus*
683 *etunicatum* (Pawlowska and Taylor 2004), although at the time the pattern was interpreted as a
684 sign of polyploidy. Incomplete lineage sorting of rDNA variants across species of
685 *Claroideoglomus* has also been attributed to the large effective population size of these
686 organisms which slow down the speed of lineage sorting and hence allows variants to be

687 maintained (Vankuren et al. 2013). This model is consistent with the relatively weak signature of
688 purifying selection observed among shared polymorphisms, as strong selection would lead to fast
689 sorting of variants over evolutionary time.
690
691 A second model is based on the idea that nuclei with variable functions can complement each
692 other within multinucleate cells, leading to an advantage of maintained polymorphisms at the
693 organism level. The effect prediction analyses of shared polymorphic sites indicated that even
694 highly-deleterious SNPs were maintained across strains and species, albeit at significantly lower
695 proportions than expected in the absence of purifying selection (table S7). This observation is
696 puzzling given that these include variants in genes with possible direct effects on nuclear
697 function (supplement B) which contradicts the efficiency of selection on individual nuclei.
698 Highly-deleterious mutations should not persist in the nuclear population due to their potential
699 impact on nuclear performance and function unless the observed polymorphism is maintained as
700 a result of complementary traits of different nuclei. Thus, it is plausible that nuclei with highly-
701 deleterious mutations (table S7; fig S14), including those that result in loss of function, could
702 cooperate with co-occurring nuclei carrying complementary genes thereby effectively buffering
703 the deleterious mutations and maintaining a functional organism.
704
705 In an alternative scenario of this second model, different nuclei carry different genes, or gene
706 variants, which contribute differently to functional traits beneficial at the organism level.
707 Asynchronized nuclear division gives AM fungi the potential for variable nucleotype frequencies
708 in different parts of the organism that experience different environments. This phenomenon was
709 suggested by Angelard et al. (2010, 2013) to account for the significant differences in

710 phenotypes observed in single spore lines of *R. irregularis* as compared to the parental lines.
711 The authors attributed the observed allelic differences to a scenario in which different alleles are
712 located in different nuclei (Angelard et al. 2010; 2013). Similarly, it has been shown that changes
713 in nucleotype frequencies within AM fungal strains affect their symbiotic performance and
714 response to different environments and hosts (Kokkoris et al. 2020; Cornell et al. 2022). In this
715 model where co-occurring nuclei complement each other, frequency-dependent selection
716 subsequently regulates the organismal allele frequencies in response to both the organism's
717 expressed phenotype and to different biotic and abiotic factors. This model also fit the data on
718 the filamentous ascomycete *Neurospora tetrasperma*, for which Meunier et al. (2018) suggested
719 that complementary nuclei co-evolve to maximize the fitness of the organism, and hence, act
720 cooperatively and facilitate adaptation by generating complementary phenotypes. Our finding of
721 intermediate frequencies of polymorphisms shared between strains and species, as opposed to
722 strain-specific SNPs that rarely reach higher frequencies, is consistent with both scenarios of the
723 second model of nuclei carrying complementary traits (fig. S12-S13).
724
725 In a third model, conflict may arise between the fitness of the nuclei and the fitness of the
726 organism, as demonstrated in *N. crassa* and *N. tetrasperma* where nuclei have been observed to
727 compete (Bastiaans, Debets, and Aanen 2016; Meunier et al. 2018; Grum-Grzhimaylo et al.
728 2021). According to this model, mutations with no effect or with positive effect on the
729 performance of the nuclei may have detrimental effects on the performance of the organism but
730 will exploit the function of the alternative nuclei, and not be purged by selection at the nuclear
731 level until they reach high frequencies. Because all nuclei in the mycelium are replicating
732 asynchronously (Jany and Pawlowska 2010), we assume that in a population of AM fungal

nuclei, allele frequencies will change as nuclei replicate and mix in the mycelium, with the "selfish" nuclei making up a larger proportion of the total nuclear population, potentially leading to reduced organism fitness. Frequency-dependent selection can act on the organism at different stages of the AM fungal life-cycle such as spore germination, root colonization, and symbiotic performance.

Together, the latter two models described above invoke selection at the level of the nuclei to explain the maintenance of polymorphism over evolutionary time. In our study, we also detected strong purifying selection in the fixed differences between the two AM fungal species. The overall observation of selection acting on nuclei and on the organism is congruent with findings from studies that have quantified and reported multi-level selection within and between individuals in modular organisms such as in the clonal seagrass *Zostera marina* (Yu et al. 2020), and in the filamentous fungus *N. tetrasperma* (Meunier et al. 2018). Overall, our findings provide further evidence for the hypothesis that selection acts on different levels in AM fungi.

Conclusion

In general, teasing apart the evolutionary processes driving evolution in any organism is a daunting task, and establishing the possible mechanisms that shape genetic variation within and between species in the multinucleate coenocytic AM fungi is even more challenging. Empirical studies exploring the possible mechanisms involved in maintenance of polymorphism in AM fungi are still scarce. With our unique hierarchical high-quality genomic dataset of nuclei, strains and species of AM fungi, we are able to unravel purifying selection on nuclei as an important force in their life history. This is particularly significant since these organisms are coenocytic,

haploid, and reproduce using multinucleated propagules. Our findings indicate that multi-level selection plays a potential role in the maintenance of polymorphism while limiting the population mutation load within and across AM fungal strains. Furthermore, the presence of ancestral polymorphism shared across AM fungal species indicates that the studied species evolved as diverging populations of nuclei. To further explore the extent of shared polymorphism in AM fungi and to acquire a comprehensive understanding of the forces shaping their diversity and evolution, future studies should include a more extensive sampling of species and strains. More studies of single nuclei genetics are also necessary to draw conclusions on the generality of these findings across divergent lineages of AM fungi.

Data accessibility

Raw reads are available at ENA with the project numbers PRJEB45340 and PRJNA528883.

Authors contribution statement

The study design was conceptualized by M.S-G., M.M-N., H.J., J.B., and A.R.. M.M-N generated the single nuclei data for an earlier paper and assembled the single nuclei genomes with M.S-G. and G.C.. M.S-G. and G.C. screened assemblies for the mat locus and performed mat locus PCRs. D.M. performed mapping, SNP calling, and all downstream analysis and data visualization with support from M.S-G. and M.H.T.. The manuscript was written by D.M., M.H.T., H.J., and A.R. with feedback and approval from all authors.

Funding

This study was funded by the European Research Council under the project number 678792.

779 **Acknowledgements**

780 Bioinformatic analysis and data release was possible thanks to the National Bioinformatics
781 Infrastructure Sweden (NBIS) and enabled by resources provided by the Swedish National
782 Infrastructure for Computing (SNIC), partially funded by the Swedish Research Council through
783 grant agreement no. 2018-05973. Funding for this project was provided by ERC (678792). We
784 acknowledge valuable discussions and suggestions from Ben Auxier, and Estelle Proux-Wéra at
785 the National Bioinformatics Infrastructure Sweden (NBIS) at SciLifeLab for bioinformatics
786 advice.

787

Literature Cited

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Angelard, Caroline, Alexandre Colard, Hélne Niculita-Hirzel, Daniel Croll, and Ian R. Sanders. 2010. "Segregation in a Mycorrhizal Fungus Alters Rice Growth and Symbiosis-Specific Gene Transcription." *Current Biology* 20 (13): 1216–21. https://doi.org/10.1016/j.cub.2010.05.031.

Angelard, Caroline, Colby J. Tanner, Pierre Fontanillas, Hélène Niculita-Hirzel, Frédéric Masclaux, and Ian R. Sanders. 2013. "Rapid Genotypic Change and Plasticity in Arbuscular Mycorrhizal Fungi Is Caused by a Host Shift and Enhanced by Segregation." *The ISME Journal 2014 8:2* 8 (2): 284–94. https://doi.org/10.1038/ismej.2013.154.

Auxier, Benjamin, and Anna Bazzicalupo. 2019. "Comment on 'single Nucleus Sequencing Reveals Evidence of Inter-Nucleus Recombination in Arbuscular Mycorrhizal Fungi'." *ELife* 8 (October). https://doi.org/10.7554/ELIFE.47301.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Https://Home.Liebertpub.Com/Cmb* 19 (5): 455–77. https://doi.org/10.1089/CMB.2012.0021.

Bastiaans, Eric, Alfons J.M. Debets, and Duur K. Aanen. 2016. "Experimental Evolution Reveals That High Relatedness Protects Multicellular Cooperation from Cheaters." *Nature Communications 2016 7:1* 7 (1): 1–10. https://doi.org/10.1038/ncomms11435.

811 Bécard, G., and P. E. Pfeffer. 1993. "Status of Nuclear Division in Arbuscular Mycorrhizal
812 Fungi during in Vitro Development." *Protoplasma 1993 174:1* 174 (1): 62–68.
813 https://doi.org/10.1007/BF01404043.
814 Bever, James D., Joseph B. Morton, Janis Antonovics, and Peggy A. Schultz. 1996. "Host-
815 Dependent Sporulation and Species Diversity of Arbuscular Mycorrhizal Fungi in a Mown
816 Grassland." *The Journal of Ecology* 84 (1): 71. https://doi.org/10.2307/2261701.
817 Bever, James D., and Mei Wang. 2005. "Hyphal Fusion and Multigenomic Structure." *Nature*
818 *2005 433:7022* 433 (7022): E3–4. https://doi.org/10.1038/nature03294.
819 Bonfante, Paola, and Andrea Genre. 2010. "Mechanisms Underlying Beneficial Plant–Fungus
820 Interactions in Mycorrhizal Symbiosis." *Nature Communications 2010 1:1* 1 (1): 1–11.
821 https://doi.org/10.1038/ncomms1046.
822 Boon, Eva, Sébastien Halary, Eric Bapteste, and Mohamed Hijri. 2015. "Studying Genome
823 Heterogeneity within the Arbuscular Mycorrhizal Fungal Cytoplasm." *Genome Biology and*
824 *Evolution* 7 (2): 505–21. https://doi.org/10.1093/gbe/evv002.
825 Bushnell, Author, Brian Bushnell, Rob Egan, Alex Copeland, Brian Foster, Alicia Clum, Hui
826 Sun, et al. 2014. "Mapping Single Molecule Sequencing Reads Using Basic Local
827 Alignment with Successive Refinement (BLASR): Application and Theory." *BMC*
828 *Bioinformatics* 13 (1). https://doi.org/10.1186/1471-2105-13-238.
829 Chen, Eric C.H., Stephanie Mathieu, Anne Hoffrichter, Kinga Sedzielewska-Toro, Max Peart,
830 Adrian Pelin, Steve Ndikumana, et al. 2018. "Single Nucleus Sequencing Reveals Evidence
831 of Inter-Nucleus Recombination in Arbuscular Mycorrhizal Fungi." *ELife* 7: 1–17.
832 https://doi.org/10.7554/eLife.39813.
833 Chen, Eric C.H., Emmanuelle Morin, Denis Beaudet, Jessica Noel, Gokalp Yildirir, Steve

834 Ndikumana, Philippe Charron, et al. 2018. “High Intraspecific Genome Diversity in the
835 Model Arbuscular Mycorrhizal Symbiont Rhizophagus Irregularis.” *New Phytologist* 220
836 (4): 1161–71. https://doi.org/10.1111/nph.14989.
837 Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan
838 J. Land, Xiangyi Lu, and Douglas M. Ruden. 2012. “A Program for Annotating and
839 Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome
840 of Drosophila Melanogaster Strain W1118; Iso-2; Iso-3.” *Fly* 6 (2): 80–92.
841 https://doi.org/10.4161/FLY.19695/SUPPL_FILE/KFLY_A_10919695_SM0001.ZIP.
842 Cornell, Calvin, Vasilis Kokkoris, Bianca Turcu, Jeremy Dettman, Franck Stefani, and Nicolas
843 Corradi. 2022. “The Arbuscular Mycorrhizal Fungus Rhizophagus Irregularis Harmonizes
844 Nuclear Dynamics in the Presence of Distinct Abiotic Factors.” *Fungal Genetics and*
845 *Biology* 158 (January): 103639. https://doi.org/10.1016/J.FGB.2021.103639.
846 Croll, Daniel, and Ian R. Sanders. 2009. “Recombination in Glomus Intraradices, a Supposed
847 Ancient Asexual Arbuscular Mycorrhizal Fungus.” *BMC Evolutionary Biology* 9 (1): 1–11.
848 https://doi.org/10.1186/1471-2148-9-13/FIGURES/2.
849 Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A.
850 DePristo, Robert E. Handsaker, et al. 2011. “The Variant Call Format and VCFtools.”
851 *Bioinformatics* 27 (15): 2156–58. https://doi.org/10.1093/bioinformatics/btr330.
852 Danecek, Petr, and Shane A McCarthy. 2017. “BCFtools/Csq: Haplotype-Aware Variant
853 Consequences.” *Bioinformatics* 33 (13): 2037–39.
854 https://doi.org/10.1093/BIOINFORMATICS/BTX100.
855 Furrazola, Eduardo, Ricardo Herrera-Peraza, Wittaya Kaonongbua, and James D. Bever. 2010.
856 “Glomus Candidum, a New Species of Arbuscular Mycorrhizal Fungi from North American

857 Grassland." *Mycotaxon* 113 (July): 101–9. https://doi.org/10.5248/113.101.

858 Gandolfi, Andrea, Ian R. Sanders, Valeria Rossi, and Paolo Menozzi. 2003. "Evidence of

859 Recombination in Putative Ancient Asexuals." *Molecular Biology and Evolution* 20 (5):

860 754–61. https://doi.org/10.1093/MOLBEV/MSG081.

861 Gill, D. E., Chao Lin, S. L. Perkins, and J. B. Wolf. 1995. "GENETIC MOSAICISM IN

862 PLANTS AND CLONAL ANIMALS."

863 *Https://Doi.Org/10.1146/Annurev.Es.26.110195.002231* 26 (November): 423–44.

864 https://doi.org/10.1146/ANNUREV.ES.26.110195.002231.

865 Giovannetti, Manuela, Dario Azzolini, and Anna Silvia Citernesi. 1999. "Anastomosis

866 Formation and Nuclear and Protoplasmic Exchange in Arbuscular Mycorrhizal Fungi."

867 *Applied and Environmental Microbiology* 65 (12): 5571–75.

868 https://doi.org/10.1128/aem.65.12.5571-5575.1999.

869 Grum-Grzhimaylo, Alexey A., Eric Bastiaans, Joost van den Heuvel, Cristina Berenguer

870 Millanes, Alfons J.M. Debets, and Duur K. Aanen. 2021. "Somatic Deficiency Causes

871 Reproductive Parasitism in a Fungus." *Nature Communications 2021 12:1* 12 (1): 1–9.

872 https://doi.org/10.1038/s41467-021-21050-5.

873 Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. 2013. "QUAST:

874 Quality Assessment Tool for Genome Assemblies." *Bioinformatics* 29 (8): 1072–75.

875 https://doi.org/10.1093/BIOINFORMATICS/BTT086.

876 Hijri, Mohamed, and Ian R. Sanders. 2005. "Low Gene Copy Number Shows That Arbuscular

877 Mycorrhizal Fungi Inherit Genetically Different Nuclei." *Nature* 433 (7022): 160–63.

878 https://doi.org/10.1038/nature03069.

879 House, Geoffrey L., Saliya Ekanayake, Yang Ruan, Ursel M.E. Schütte, Wittaya Kaonongbua,

880 Geoffrey Fox, Yuzhen Ye, and James D. Bevera. 2016. “Phylogenetically Structured
881 Differences in RRNA Gene Sequence Variation among Species of Arbuscular Mycorrhizal
882 Fungi and Their Implications for Sequence Clustering.” *Applied and Environmental*
883 *Microbiology* 82 (16): 4921–30. https://doi.org/10.1128/AEM.00816-
884 16/SUPPL_FILE/ZAM999117331SO1.PDF.
885 Huson, Daniel H., and David Bryant. 2006. “Application of Phylogenetic Networks in
886 Evolutionary Studies.” *Molecular Biology and Evolution* 23 (2): 254–67.
887 https://doi.org/10.1093/MOLBEV/MSJ030.
888 Jany, Jean Luc, and Teresa E. Pawlowska. 2010. “Multinucleate Spores Contribute to
889 Evolutionary Longevity of Asexual Glomeromycota.” *American Naturalist* 175 (4): 424–
890 35. https://doi.org/10.1086/650725.
891 Jinks, J. L. 1952. “Heterokaryosis: A System of Adaptation in Wild Fungi.” *Proceedings of the*
892 *Royal Society of London. Series B - Biological Sciences* 140 (898): 83–99.
893 https://doi.org/10.1098/RSPB.1952.0046.
894 Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment
895 Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and*
896 *Evolution* 30 (4): 772–80. https://doi.org/10.1093/MOLBEV/MST010.
897 Knaus, Brian J., and Niklaus J. Grünwald. 2017. “Vcfr: A Package to Manipulate and Visualize
898 Variant Call Format Data in R.” *Molecular Ecology Resources* 17 (1): 44–53.
899 https://doi.org/10.1111/1755-0998.12549.
900 Kokkoris, Vasilis, Franck Stefani, Yolande Dalpé, Jeremy Dettman, and Nicolas Corradi. 2020.
901 “Nuclear Dynamics in the Arbuscular Mycorrhizal Fungi.” *Trends in Plant Science* 25 (8):
902 765–78. https://doi.org/10.1016/J.TPLANTS.2020.05.002.

903 Kryazhimskiy, Sergey, and Joshua B. Plotkin. 2008. “The Population Genetics of DN/DS.”
904 *PLOS Genetics* 4 (12): e1000304. https://doi.org/10.1371/JOURNAL.PGEN.1000304.
905 Li, H., and R. Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler
906 Transform.” *Bioinformatics* 25 (14): 1754–60.
907 https://doi.org/10.1093/bioinformatics/btp324.
908 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.
909 Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence
910 Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.
911 https://doi.org/10.1093/bioinformatics/btp352.
912 Lin, Kui, Erik Limpens, Zhonghua Zhang, Sergey Ivanov, Diane G. O. Saunders, Desheng Mu,
913 Erli Pang, et al. 2014. “Single Nucleus Genome Sequencing Reveals High Similarity among
914 Nuclei of an Endomycorrhizal Fungus.” *PLOS Genetics* 10 (1): e1004078.
915 https://doi.org/10.1371/JOURNAL.PGEN.1004078.
916 Marleau, Julie, Yolande Dalpé, Marc St-Arnaud, and Mohamed Hijri. 2011. “Spore
917 Development and Nuclear Inheritance in Arbuscular Mycorrhizal Fungi.” *BMC*
918 *Evolutionary Biology 2011 11:1* 11 (1): 1–11. https://doi.org/10.1186/1471-2148-11-51.
919 Mateus, Ivan D., Ben Auxier, Mam M.S. Ndiaye, Joaquim Cruz, Soon Jae Lee, and Ian R.
920 Sanders. 2022. “Reciprocal Recombination Genomic Signatures in the Symbiotic
921 Arbuscular Mycorrhizal Fungi Rhizophagus Irregularis.” *PLOS ONE* 17 (7): e0270481.
922 https://doi.org/10.1371/JOURNAL.PONE.0270481.
923 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, et
924 al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-
925 Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303.

926 https://doi.org/10.1101/gr.107524.110.

927 Meunier, Cécile, Sara Hosseini, Nahid Heidari, Zaywa Maryush, and Hanna Johannesson. 2018.
928 "Multilevel Selection in the Filamentous Ascomycete Neurospora Tetrasperma." *American*
929 *Naturalist* 191 (3): 290–305. https://doi.org/10.1086/695803.

930 Montoliu-Nerin, Merce, Marisol Sánchez-García, Claudia Bergin, Manfred Grabherr, Barbara
931 Ellis, Verena Esther Kutschera, Marcin Kierczak, Hanna Johannesson, and Anna Rosling.
932 2020. "Building de Novo Reference Genome Assemblies of Complex Eukaryotic
933 Microorganisms from Single Nuclei." *Scientific Reports 2020 10:1* 10 (1): 1–10.
934 https://doi.org/10.1038/s41598-020-58025-3.

935 Montoliu-Nerin, Merce, Marisol Sánchez-García, Claudia Bergin, Verena Esther Kutschera,
936 Hanna Johannesson, James D. Bever, and Anna Rosling. 2021. "In-Depth Phylogenomic
937 Analysis of Arbuscular Mycorrhizal Fungi Based on a Comprehensive Set of de Novo
938 Genome Assemblies." *Frontiers in Fungal Biology* 2 (September): 53.
939 https://doi.org/10.3389/ffunb.2021.716385.

940 Morin, Emmanuelle, Shingo Miyauchi, Hélène San Clemente, Eric C. H. Chen, Adrian Pelin,
941 Ivan de la Providencia, Steve Ndikumana, et al. 2019. "Comparative Genomics of
942 Rhizophagus Irregularis, R. Cerebriforme, R. Diaphanus and Gigaspora Rosea Highlights
943 Specific Genetic Features in Glomeromycotina." *New Phytologist* 222 (3): 1584–98.
944 https://doi.org/10.1111/NPH.15687.

945 Ortiz, Edgardo M. 2019. "Vcf2phylip v2.0: Convert a VCF Matrix into Several Matrix Formats
946 for Phylogenetic Analysis.," January. https://doi.org/10.5281/ZENODO.2540861.

947 Otto, S. P., and M. E. Orive. 1995. "Evolutionary Consequences of Mutation and Selection
948 within an Individual." *Genetics* 141 (3): 1173–87.

949 https://doi.org/10.1093/genetics/141.3.1173.

950 Palmer, Jon, and Jason Stajich. 2020. “Nextgenusfs/Funannotate: Funannotate v1.8.9,” March.

951 https://doi.org/10.5281/ZENODO.2604804.

952 Pawlowska, Teresa E., and John W. Taylor. 2004. “Organization of Genetic Variation in

953 Individuals of Arbuscular Mycorrhizal Fungi.” *Nature 2004 427:6976* 427 (6976): 733–37.

954 https://doi.org/10.1038/nature02290.

955 Pineda-Krch, M., and K. Lehtilä. 2004. “Costs and Benefits of Genetic Heterogeneity within

956 Organisms.” *Journal of Evolutionary Biology* 17 (6): 1167–77.

957 https://doi.org/10.1111/J.1420-9101.2004.00808.X.

958 Quinlan, Aaron R. 2014. “BEDTools: The Swiss-Army Tool for Genome Feature Analysis.”

959 *Current Protocols in Bioinformatics* 47 (1): 11.12.1-11.12.34.

960 https://doi.org/10.1002/0471250953.BI1112S47.

961 R Core Team. 2021. “R: A Language and Environment for Statistical Computing. R Foundation

962 for Statistical Computing, Vienna, Austria.” https://www.r-project.org/.

963 Redecker, D., R. Kodner, and L. E. Graham. 2000. “Glomalean Fungi from the Ordovician.”

964 *Science* 289 (5486): 1920–21. https://doi.org/10.1126/science.289.5486.1920.

965 Richard, Michod. 1996. “Cooperation and Conflict in the Evolution of Individuality. II. Conflict

966 Mediation.” *Proceedings: Biological Sciences 263, No. 1372*, 813–22.

967 http://www.jstor.org/sTable/50560.

968 Ropars, Jeanne, Kinga Sędzielewska Toro, Jessica Noel, Adrian Pelin, Philippe Charron, Laurent

969 Farinelli, Timea Marton, et al. 2016. “Evidence for the Sexual Origin of Heterokaryosis in

970 Arbuscular Mycorrhizal Fungi.” *Nature Microbiology 2016 1:6* 1 (6): 1–9.

971 https://doi.org/10.1038/nmicrobiol.2016.33.

Sanders, Ian R., and Daniel Croll. 2010. "Arbuscular Mycorrhiza: The Challenge to Understand the Genetics of the Fungal Partner." *Http://Dx.Doi.Org/10.1146/Annurev-Genet-102108-134239* 44 (November): 271–92. https://doi.org/10.1146/ANNUREV-GENET-102108-134239.

Santelices, Bernabé. 1999. "How Many Kinds of Individual Are There?" *Trends in Ecology & Evolution* 14 (4): 152–55. https://doi.org/10.1016/S0169-5347(98)01519-5.

Sędzielewska, Kinga A., Jörg Fuchs, Eva M. Temsch, Keith Baronian, Roland Watzke, and Gotthard Kunze. 2011. "Estimation of the Glomus Intraradices Nuclear DNA Content." *New Phytologist* 192 (4): 794–97. https://doi.org/10.1111/J.1469-8137.2011.03937.X.

Serghi, Edward Umberto, Vasilis Kokkoris, Calvin Cornell, Jeremy Dettman, Franck Stefani, and Nicolas Corradi. 2021. "Homo- and Dikaryons of the Arbuscular Mycorrhizal Fungus Rhizophagus Irregularis Differ in Life History Strategy." *Frontiers in Plant Science* 12 (August): 1544. https://doi.org/10.3389/FPLS.2021.715377/BIBTEX.

Smit, A.F, and R Hubley. 2008. "RepeatModeler Open-1.0." http://www.repeatmasker.org.

Smit, A.F, R Hubley, and P Green. 2015. "RepeatMasker Open-4.0."

Talukdar, Narayan C., and James J. Germida. 2011. "Occurrence and Isolation of Vesicular–Arbuscular Mycorrhizae in Cropped Field Soils of Saskatchewan, Canada." *Https://Doi.Org/10.1139/M93-082* 39 (6): 567–75. https://doi.org/10.1139/M93-082.

Ter-Hovhannisyan, Vardges, Alexandre Lomsadze, Yury O. Chernoff, and Mark Borodovsky. 2008. "Gene Prediction in Novel Fungal Genomes Using an Ab Initio Algorithm with Unsupervised Training." *Genome Research* 18 (12): 1979–90. https://doi.org/10.1101/GR.081612.108.

Tisserant, Emilie, Mathilde Malbreil, Alan Kuo, Annegret Kohler, Aikaterini Symeonidi,

995 Raffaella Balestrini, Philippe Charron, et al. 2013. “Genome of an Arbuscular Mycorrhizal
996 Fungus Provides Insight into the Oldest Plant Symbiosis.” *Proceedings of the National*
997 *Academy of Sciences of the United States of America* 110 (50): 20117–22.
998 https://doi.org/10.1073/pnas.1313452110.
999 Vandenkoornhuyse, P., C. Leyval, and I. Bonnin. 2001. “High Genetic Diversity in Arbuscular
1000 Mycorrhizal Fungi: Evidence for Recombination Events.” *Heredity 2001 87:2* 87 (2): 243–
1001 53. https://doi.org/10.1046/j.1365-2540.2001.00941.x.
1002 Vankuren, Nicholas W., Henk C. den Bakker, Joseph B. Morton, and Teresa E. Pawlowska.
1003 2013. “RIBOSOMAL RNA GENE DIVERSITY, EFFECTIVE POPULATION SIZE, AND
1004 EVOLUTIONARY LONGEVITY IN ASEXUAL GLOMEROMYCOTA.” *Evolution* 67
1005 (1): 207–24. https://doi.org/10.1111/J.1558-5646.2012.01747.X.
1006 Waterhouse, Robert M, Mathieu Seppey, Felipe A Simão, Mosè Manni, Panagiotis Ioannidis,
1007 Guennadi Klioutchnikov, Evgenia V Kriventseva, and Evgeny M Zdobnov. 2018. “BUSCO
1008 Applications from Quality Assessments to Gene Prediction and Phylogenomics.” *Molecular*
1009 *Biology and Evolution* 35 (3): 543–48. https://doi.org/10.1093/MOLBEV/MSX319.
1010 Weir, B. S., and C. Clark Cockerham. 1984. “Estimating F-Statistics for the Analysis of
1011 Population Structure.” *Evolution* 38 (6): 1358. https://doi.org/10.2307/2408641.
1012 Wyss, Tania, Frédéric G Masclaux, Pawel Rosikiewicz, Marco Pagni, and Ian R Sanders. 2016.
1013 “Population Genomics Reveals That Within-Fungus Polymorphism Is Common and
1014 Maintained in Populations of the Mycorrhizal Fungus Rhizophagus Irregularis.” *The ISME*
1015 *Journal 2016 10:10* 10 (10): 2514–26. https://doi.org/10.1038/ismej.2016.29.
1016 Ye, Jian, George Coulouris, Irena Zaretskaya, Ioana Cutcutache, Steve Rozen, and Thomas L
1017 Madden. 2012. “Primer-BLAST: A Tool to Design Target-Specific Primers for Polymerase

1018 Chain Reaction." *BMC Bioinformatics 2012 13:1* 13 (1): 1–11.
1019 https://doi.org/10.1186/1471-2105-13-134.
1020 Yildirir, Gökalp, Jana Sperschneider, Mathu Malar C, Eric C.H. Chen, Wataru Iwasaki, Calvin
1021 Cornell, and Nicolas Corradi. 2022. "Long Reads and Hi-C Sequencing Illuminate the Two-
1022 Compartment Genome of the Model Arbuscular Mycorrhizal Symbiont Rhizophagus
1023 Irregularis." *New Phytologist* 233 (3): 1097–1107. https://doi.org/10.1111/NPH.17842.
1024 Yu, Lei, Christoffer Boström, Sören Franzenburg, Till Bayer, Tal Dagan, and Thorsten B.H.
1025 Reusch. 2020. "Somatic Genetic Drift and Multilevel Selection in a Clonal Seagrass."
1026 *Nature Ecology and Evolution* 4 (7): 952–62. https://doi.org/10.1038/s41559-020-1196-4.
1027