Submission type: **Research article**

Recombinant haplotypes reflect sexual reproduction in symbiotic arbuscular mycorrhizal fungi

Ivan D. Mateus[1]*, Ben Auxier[2], Mam M. S. Ndiaye[1], Joaquim Cruz[1], Soon-Jae Lee[1] & Ian R. Sanders[1]

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

[2]*Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands.*

* Author for correspondence: ivandario.mateusgonzalez@unil.ch

Keywords: Arbuscular mycorrhizal fungi, AMF, haplotypes, recombination, sexual reproduction

Abstract

Arbuscular mycorrhizal fungi (AMF) are part of the most widespread fungal-plant symbiosis. They colonize at least 80% of plant species, promote plant growth and plant diversity. These fungi are multinucleated and display either one or two nucleus genotypes (monokaryon and dikaryon) determined by a putative mating-type locus. This taxon has been considered as an ancient asexual scandal because of the lack of observable sexual structures. Despite the identification of a putative mating-type (MAT-type) locus and the functional activation of genes related to mating when two isolates co-exist, it is still unknown if AMF display a sexual or a parasexual life cycle.

To test if AMF genomes display signatures of a sexual life-cycle involving the putative MAT-locus, we used publicly available genome sequences to test if recombining nucleotype-specific haplotypes could be identified using short-read Illumina sequences. We identified nucleus genotype-specific haplotypes within dikaryons and compared them to orthologous gene sequences from related monokaryon isolates displaying similar putative MAT-types. We show that haplotypes within a dikaryon isolate are more similar to homologue sequences of isolates having the same MAT-type than among them. We demonstrate that these genotype-specific haplotypes are recombinant, and are not consistently most similar to the monokaryon isolate sharing the same mating-type allele.

These results are consistent with a sexual origin of the dikaryon rather than a parasexual origin and provides an important step to understand the life cycle of these globally important symbiotic fungi.

Introduction

Arbuscular mycorrhizal fungi are plant symbionts, forming symbioses with most plant species, promoting plant growth (Harrison, 1997), plant community diversity (van der Heijden *et al.*, 1998; Antunes *et al.*, 2011) and how plants cope with biotic (Thygesen *et al.*, 2004) or abiotic stresses (Augé, 2001). As a consequence, they are widely used in agriculture (Smith & Read, 2008). Within isolate genetic variability in AMF has been reported to result in differential effects on plant growth (Angelard *et al.*, 2010). Understanding how genetic variability is generated in AMF, is important because it could be harnessed to generate genetic variants that could be beneficial for their plant hosts (Sanders, 2010).

AMF are part of the Glomeromycotina subphylum (Spatafora *et al.*, 2016), which fossil records date to at least ~400 Million years ago (Remy *et al.*, 1994). They are coenocytic (without septa separating otherwise adjacent compartments), their hyphae harbor hundreds of nuclei within the same cytoplasm (Marleau *et al.*, 2011) and no single-nucleus state has ever been recorded in these taxa. The nuclei of these fungi have been reported as haploid (Ropars *et al.*, 2016; Kobayashi *et al.*, 2018; Morin *et al.*, 2019b). This group of fungi has been previously considered as an ancient asexual scandal (Judson & Normark, 1996), because of low morphological diversification and the absence of observable sexual structures leading to the conclusion that they only reproduce asexually. However, evidence suggests that sexual reproduction could be possible in AMF because these fungi contain a complete meiosis machinery (Halary *et al.*, 2011). Furthermore, a putative MAT-locus has been proposed (Ropars *et al.*, 2016), population genetic data suggests the existence of recombination in AMF populations (Croll & Sanders, 2009) and activation of genes related to mating has been detected when different isolates of the same species co-exist in plant roots (Mateus *et al.*, 2020). Inter-nucleus recombination has been reported (Chen *et al.*, 2018a), although the robustness of the analysis has been questioned. The application of strict filtering parameters such as removal of heterozygous sites in haploid nuclei, duplicated regions of the genome, and low-coverage depths base calls results in an extreme loss of

recombination signal (Auxier & Bazzicalupo, 2019), calling the conclusions of the study into question. Although some of these limitations, as coverage depth and filtering-out heterozygous sites were addressed (Chen *et al.*, 2020), other limitations such as replicability (recombination events shown in several nuclei sharing the same putative MAT-type) and issues related to the whole-genome amplification step as the formation of chimeric sequences (Yilmaz & Singh, 2012), allelic drop-out (Lauri *et al.*, 2013) and SNP miscalling (Ning *et al.*, 2014) are inherent limitations of the analysis of single nucleus amplification data.

Like most fungi, AMF can undergo anastomosis, the fusion of hyphae. Through these connections, bi-directional flow of cytoplasm has been observed between genetically different AMF individuals (Giovannetti *et al.*, 1999). Via anastomosis, the transfer of genetic material (parasexuality) has been suggested, as well, as a mechanism of maintenance of genetic diversity in the absence of sexual recombination (Bever & Morton, 1999). However, the existence and relevance of sexuality and/or parasexuality for the evolution of AMF remains unknown (Yildirir *et al.*, 2020). It is still unknown whether a sexual event comprising meiotic recombination, or a parasexual event, could influence the transition between dikaryon and monokaryon isolates.

In the past, there was a debate about nuclei being identical (homokaryosis) or different (heterokaryosis). Today, it is widely accepted that the model AMF species *Rhizophagus irregularis* exists either as monokaryons (such as isolates DAOM197198, A1,B12,C2) or dikaryons (such as isolates A4, A5, SL1, C3 and G1) (Ropars *et al.*, 2016; Chen *et al.*, 2018a; Masclaux *et al.*, 2018; Kokkoris *et al.*, 2021). Single-nuclei from dikaryons (isolates A4, A5 and SL1) cluster into two genetically different nucleotypes (where nucleotype refers to the genotype of a nucleus), based on the identity of their mating type locus (MAT-locus) (Chen *et al.*, 2018a). This demonstrates that the presence of two copies of the MAT-locus is a reliable marker of the dikaryon state. However, the claimed evidence of large inter-nuclei recombination (Chen *et al.*, 2018a, 2020) does not fit the observations about a monokaryon – dikaryons organization. The presence of repeated inter-nuclear recombination in dikaryons

without prior crossing with other isolates, as observed in the single-nucleus genotypes shown in Chen et al. (2018, 2020) would result in heterokaryons with more than two types of nuclei. But this does not appear to be the case for *R. irregularis* (Ropars *et al.*, 2016; Chen *et al.*, 2018a; Masclaux *et al.*, 2019; Auxier & Bazzicalupo, 2019).

In *R. irregularis*, a putative life cycle comprising the formation of dikaryons from monokaryons containing compatible alleles at the MAT-locus has been proposed (Ropars *et al.*, 2016). A previous attempt to identify the origin of the dikaryon isolate A5 was made based on the hypothesis that the monokaryon isolates A1 (MAT-type 3) and C2 (MAT-type 6) were the parents of the dikaryon isolate A5 (MAT-type 3/ MAT-type 6). Ropars *et al.,* used single-nucleus data and observed that several positions in the genome did not support their hypothesis and concluded that these two isolates could not be the direct progenitors of isolate A5 (Ropars *et al.*, 2016). However, this analysis was limited by the analysis of only 10 single nucleus polymorphisms across 4 contigs, and included positions with more than one allele in a haploid single-nucleus, and that also show different alleles at a given position among single nuclei of monokaryon isolates. Consequently, we still do not know the origin of dikaryon isolates, whether they originate from the fusion of two monokaryon isolates, as in Basidiomycete fungi, and whether recombinant nucleotypes are detected in dikaryon isolates. To move beyond single polymorphisms, haplotype analysis would allow to exclude some of the limitations of previous methods. The haplotypes from dikaryon isolates can then be compared to orthologous sequences of related monokaryon isolates and would allow the identification of a sexual event, involving recombination, or a parasexual event in the absence of recombination that could explain the transition between AMF monokaryon and dikaryons.

The recent development of methods for long-read sequencing can be used to identify haplotypes and, consequently, nucleotype-specific haplotypes in fungal dikaryon isolates (Li *et al.*, 2019). An alternative exists from the reanalysis of short-read sequences to identify genome wide copy number variation. These analyses consist in obtaining the read depth, or

coverage, after mapping the reads to a genome assembly and identifying changes in coverage across the genome (Yoon *et al.*, 2009). The analysis of drop in coverage has resulted in the identification of chromosome copy number variation and segmental chromosome aneuploidies in several fungal species (Todd *et al.*, 2017). In AMF, a drop in coverage analysis was used to originally identify the putative MAT-locus (Ropars *et al.*, 2016) highlighting the potential to identify nucleotype-specific haplotypes in dikaryons with this technique.

Here, we demonstrated that AMF dikaryons display genetic recombination by analyzing nucleotype-specific haplotypes in dikaryon and monokaryon isolates. In this study we used publicly available whole genome and single-nucleus sequence data to identify nucleotype-specific haplotypes in dikaryon isolates. We identified regions displaying drops in coverage in whole genome sequence assemblies. In these regions we detected the presence of genes that have two copies in the dikaryon isolates and one copy in monokaryons. We then confirmed independently, with genome sequence data from single-nucleus, that in dikaryon isolates, different nuclei have different alleles of the previously detected genes, showing that the genes identified in this study are highly divergent alleles and are nucleotype-specific. The identification of nucleotype-specific alleles allowed us to test whether monokaryon isolates that share the same MAT-type could be the origin of dikaryon isolates and whether recombination could be detected between the nucleotype-specific haplotypes.

Materials and Methods

Source data

We used public-available sequence reads, genome assemblies and annotations of isolates A1, A4, A5, C2 of *R. irregularis* for this study, including data from whole-genome and single-nucleus sequencing (Supplementary Table 1). We downloaded the sequence reads from the sequence read archive (SRA) using the SRAtoolkit software (Leinonen *et al.*, 2011). We used for the different analysis sequence reads from whole-genome and single-nucleus sequence data.

Coverage analysis

We first trimmed the sequence reads using Trimgalore (Krueger, 2015) with the default parameters. We then used BWA (Li, 2013) to index the reference genome assemblies and BWA mem -M (Li, 2013) to map the reads to the reference whole-genome assemblies. We mapped the reads coming from a given isolate to the reference genome assembly of the same isolate (i.e. reads A1 mapped to reference A1). We then kept the reads that display a mapping quality of at least 30. We used the genomecov tool from bedtools (Quinlan & Hall, 2010) to calculate the coverage for each position. We then created a ready-to-use algorithm that detects genome-wide drop in coverage analysis in whole-genome data (Supplementary File 1). The algorithm divides the data in portions of 50000bp. Then, with a sliding window approach consisting of windows of 400bp and steps of 100bp, the algorithm searches for drops in coverage of 0.3-0.6 times lower than the median coverage of the entire genome and that display a minimum length of 1000bp (Please refer to Supplementary File 1 for the algorithm specifications). We then further filtered the drop in coverage regions by keeping

only the regions that display an average of 1.25 coverage difference between the neighboring regions and the drop in coverage region.

Gene detection in drops

We identified all the genes located within genomic regions that presented a drop of sequencing coverage. We used the 'intersect' command from the BEDTools suite with the existing gene annotations corresponding to each *R. irregularis* isolate (GTF format) and their query regions with drops in coverage (BED format) to find the overlapping genes (Quinlan & Hall, 2010). Genes in scaffolds smaller that 1kb were not considered for further analyses.

***de novo* single-nucleus assemblies**

We trimmed the raw reads by using TrimGalore-0.6.0 (Krueger, 2015) with default parameters. After trimming, we performed single-nucleus *de novo* assemblies with SPAdes v3.14 (Bankevich *et al.*, 2012) with the following parameters: -k 21,33,55,77 --sc --careful --cov-cutoff auto. The resulted single-nucleus genome assemblies were used for further analysis. The length, number of contigs and N50 value of the *de novo* assemblies was evaluated with quast- 5.1.0rc1 with default parameters (Mikheenko *et al.*, 2018).

Identification of genes in genome assemblies

To identify the position of gene sequences on the different genome assemblies, we first extracted a query sequence. We then used the console NCBI+ blast suite (Camacho *et al.*, 2009) to blast the query against the desired target. In the case of the putative MAT-locus, we used the homeodomain genes HD2 and HD1-like as query (HD2:KT946661.1, HD1-like:

KU597387 from isolate A1). For further downstream analyses, we extracted the sequences from the genome assemblies by using the blastdbcmd command from the NCBI+ suite. We used a reciprocal blast approach to identify the gene sequences corresponding between the whole genome sequence data and the single-nucleus data. We considered the bests hits by evaluating the % identity, mismatches, e-value and bitscore.

Orthology inference of genes present in drops in coverage regions

We used Orthofinder 2.3.11 (Emms & Kelly, 2019) to identify orthologs of genes that were found inside the drop in coverage regions within the same isolate. We used the orthogroups output from Orthofinder for the different analyses.

Synteny plots

We compared genomic regions by performing synteny plots computed with EasyFig2.2.3 (Sullivan *et al.*, 2011). We provide full Genbank files to compare genomic regions to each other. The software executes a blast comparison between the regions to determine their homology.

Genetic distance between nucleotypes

Coding sequences for the 12 confirmed nucleotype-specific genes were extracted using the Blast+ command line blastdbcmd tool (Camacho *et al.*, 2009). The sequences were then aligned with MAFFT (Katoh *et al.*, 2017) using the --auto option. Then, the ape package (Paradis *et al.*, 2004) of R was used to calculate the pairwise distance between the 4 alleles

(2 from A5, and 1 each from A1 and C2), and plot a distance tree of the 4 alleles from which the quartet arrangement was determined. The mean distance was then calculated for the combined set of the 12 confirmed nucleotype-specific genes.

Recombination detection

We compared the sequences from drop in coverage regions from both nucleotypes of isolate A5 and isolates A1 and C2 to detect if isolate A5 display recombination events between the two putative parental isolates. After identification of the syntenic region among the different isolates, we aligned the sequences with MAFFT (Katoh *et al.*, 2017) and evaluated whether the sequence of one of the nucleotype of isolate A5 was similar to A1 and the other similar to C2.

Phylogenetic analyses

We used MEGA-X (Kumar *et al.*, 2018) for the different phylogenetic reconstructions shown in the study. We first aligned the data with ClustalW. We then find the best DNA models describing the relation between the sequences. Finally, we used a maximum likelihood phylogeny reconstruction with 100 bootstraps to infer the phylogenetic relation among the samples. In several cases, we were not able to perform maximum likelihood phylogenies because of the low number of samples to compare, so UPGMA trees were done instead. Phylogenetic reconstructions of the different orthologous groups on Figure 4 where produced by the Orthofinder software.

Results

Drop in coverage analysis reveals potential nucleotype-specific haplotypes

Previously, a drop in coverage analysis was used for the identification of a putative MAT-locus in *R. irregularis* (Ropars *et al.*, 2016). Although the identification of drops in coverage at other loci were detected in *R. Irregularis* (Ropars *et al.*, 2016), no further description was made on those other regions. We developed a ready-to-use algorithm (Supplementary File 1) that allows us to identify genome-wide drop in coverage events (for the accessions of raw data and genome assemblies used in this study see Supplementary Table 1).

We identified drops in coverage in 4 different isolates of *R. irregularis* which are reported to be dikaryons (A4 and A5) and monokaryons (A1 and C2) (Figure 1a, Supplementary Table 2). However, the number of coverage drops was different between dikaryons (A4-A5) and monokaryons (A1-C2). Isolates A4 and A5 displayed 1145 and 1032 drops in coverage, respectively, at different loci while isolates A1 and C2 displayed 129 and 121, respectively (Figure1b). In the regions where a drop in coverage region was observed, in total we identified 499, 444,16 and 24 genes in isolates A4, A5, A1 and C2 respectively (Figure 1c, Supplementary Table 3). These results confirmed that dikaryon isolates displayed more heterozygous regions than the monokaryons. This suggests that the genes present within the regions showing a drop in coverage are potential candidates for nucleotype-specific alleles in dikaryons.

The total length of the regions where a drop in coverage occurred covered 2.37 % of isolate A4, 2.21 % of isolate A5 and 0.01% of isolate A1 and isolate C2. We identified that 9.1 %

and 6.6% of the contigs of isolates A4 and A5 contained regions where a drop in coverage was observed, while they were very rare (1%) in genome assemblies of isolates A1 and C2 (Figure 1d). These results suggest that the events detected in monokaryon isolates could represent *de novo* mutations but the most likely explanation is that they could be residual technical artifacts. The length of the majority of drops in coverage regions was between 1-2kb and very few spanned more than 10kb (Figure 1d). Confirming the reliability of our approach, we detected the expected drop in coverage in the MAT-locus region in isolates A4 and A5 but not in isolates A1 and C2 (Figure 1d).

One cause for a drop in coverage could be copy number variation between the nucleotypes in a dikaryon. To test for this, we inferred orthologous gene families among the different isolates to identify if the genes present in the drop in coverage regions displayed more than one copy in their own genome. We used the gene annotation available for each isolate and performed the orthology inference on all the genes present in each genome. The orthologous inference resulted in many orthologous groups displaying more than one copy within each isolate (A4: 20%, A5: 18%, A1: 17% and C2: 19%; Supplementary figure 1, Supplementary Table 4) confirming the high incidence of paralogs in these fungi (Morin *et al.*, 2019a). We further identified orthologous gene families of genes detected in drop in coverage in isolates A4 and A5 independently. Under the assumption of a monokaryon-dikaryon genome organization in *R. irregularis*, to avoid the confounding effect of duplications and reduce the complexity of the dataset, we kept only the orthologous groups that are present in the drop in coverage regions and that display two copies in the dikaryon isolates (A4, A5) and a single copy in the monokaryons (A1, C2) (Figure 2a, Supplementary Table 5). We identified 32 orthologous groups that are present with two copies in isolate A4 and only a single copy in isolates A1 and C2. We also identified 27 orthologous groups in isolate A5 that display two copies. Only two orthologous groups were common between the two isolates: namely, HD2 and HD1-like which are part of the putative MAT-locus in *R. irregularis* (Figure 2b).

As reported in Ropars *et al.,* we observed that the two copies of the putative MAT-locus in the dikaryons were located in different contigs. One copy of HD2 and HD1-like genes were present in a long contig of the genome assembly, while the second copy was present in a much shorter contig (Figure 2c). We observed the same pattern for the other orthologous groups, where the second copy was always present in a second shorter contig (for several examples see Figure 2d). We then tested if the orthologous genes from the different isolates are orthologs and not paralogs. We performed a synteny analysis to compare the genomic location among isolates of the orthologous genes. We identified that 12 out of 32 predicted orthologs in A4 and 16 out of 27 predicted orthologs in A5, were located in the same genomic location on the different isolates, suggesting that they could be considered as orthologs (Supplementary Table 5, for examples of inferred orthologs and paralogs see Supplementary Figure 2).

Taken together, this shows that in the whole genome assemblies of dikaryon isolates, two divergent alleles were assembled into different scaffolds; one longer scaffold containing neighboring regions and a shorter scaffold without the neighboring regions. Given the haploid nuclei of the dikaryon isolates, two possibilities are consistent with this previous fact: The two copies could be present within the same nucleus or in different nuclei.

Drop in coverage signatures represent nucleotype-specific haplotypes

To confirm that genes found inside drop in coverage regions are nucleotype-specific, we used raw reads generated from whole genome sequencing of individual nuclei of dikaryon isolates A4 and A5 (Chen *et al.*, 2018a) to produce *de novo* single-nucleus assemblies. The *de novo* assemblies were very fragmented and incomplete (Supplementary table 6) and their utilization was highly limited. This limitation resulted in the inability to identify some genes

and some complete gene sequences. However, a reciprocal blast approach between the whole genome assembly and the single-nucleus assemblies allowed us to detect sequences in the single-nucleus assemblies corresponding to the genes detected in the whole genome assemblies.

We tested in the dikaryons if the genes identified in the drop in coverage regions were present in the form of different alleles in different single nuclei by using a reciprocal blast approach (Supplementary Table 7). We confirmed 9 orthologous genes to be nucleotype-specific in isolate A4 and 12 orthologous genes in isolate A5 (Table 1). We found that the population of nuclei clustered in two groups that corresponded to the identity of the MAT-locus contained in each nucleus (Figure 3). This result confirms that nucleotype-specific alleles in dikaryon isolates can be identified based on genes found in drop in coverage regions and that are represented by a duplication within the genome assembly.

Nucleotype-specific alleles from A5 share a more recent evolutionary origin with monokaryon isolates A1 and C2 than among them

The origin of dikaryon isolates could be investigated through comparisons of monokaryon isolates that display the same putative MAT-locus as those found in the dikaryons (Isolates A5:MAT-3/MAT-6; A1:MAT-3; C2: MAT-6). A phylogenetic reconstruction of the putative MAT-locus suggests that MAT-3 from isolates A1 and A5 are more closely related than MAT-6 from isolates C2 and A5 (Ropars *et al.*, 2016). Furthermore, genome-wide reduced genome representation phylogenetic reconstructions of several *R. irregularis* isolates indicated that isolate A5 is more closely related to isolate A1 than to isolate C2 (Wyss *et al.*, 2016; Savary *et al.*, 2018)

To confirm the previous findings, for each previously defined nucleotype-specific gene, we compared the phylogenetic relationship of the two nucleotype-specific alleles in isolate A5

isolate and in isolates A1, C2 and A4. We observed that for several nucleotype-specific genes, genes from isolate A1 clustered with one of the alleles of isolate A5, but it was not always the case in isolate C2 (Figure 4a).

We then analyzed the genetic distance of each of the 12 nucleotype-specific genes independently between the two alleles of isolate A5 and the homologous allele in isolates A1 and C2. The mean nucleotide distance between the two A5 alleles was 0.147. The mean distance between A1 and C2 was 0.132. In contrast, the mean of the minimum distance between an allele of isolate A5 and isolate A1 was 0.013 and between A5 and C2 was 0.043. If we randomly select one allele for each gene, into a set “a”, the mean distance for the 12 nucleotype-specific alleles for one allele from A5 (allele “a”) to A1 was 0.031, and 0.122 for the second allele (allele "b"). The distance of the same allele "a" from isolate A5 to C2 was 0.136, and the distance of the second allele "b" to C2 was 0.062 (Figure 4b). We did not observe any case where the two A5 alleles clustered together, instead we observed that for all 12 nucleotype-specific gene the two A5 alleles were more similar to the allele from isolate A1 or C2 (Figure 4c). The mean distances calculated between alleles in this study are much higher than average distances calculated on the whole genome between different isolates (Chen *et al.*, 2018b), reflecting our selection criteria for nucleotype-specific regions. As each of the A5 alleles was closer to A1 or C2, instead of the two A5 alleles being most similar, this indicates that the alleles of A5 share a more recent evolutionary origin with these monokaryons than the two alleles within A5.

Recombination between nucleotype-specific haplotypes in isolate A5

Knowledge about nucleotype-specific haplotypes of dikaryon isolate A5 and their orthologs in isolates A1 and C2 allowed us to test whether recombination occurred in nucleotype-

specific haplotypes of dikaryon isolate A5 (Figure 5a). We scanned the different nucleotype-specific-haplotypes for the detection of recombination events within the two haplotypes of isolate A5. Comparison of haplotypes from isolates A1, C2 and the two haplotypes of A5 showed that each nucleotype-specific sequence from isolate A5 was highly similar to either isolate A1 or C2, but we did not identify any recombination events within haplotypes (Figure 5b).

To further assess the potential for clonal relationships between the two nucleotypes within A5 and isolates A1 and C2, we compared the nucleotype-specific haplotypes on the single-nucleus assemblies from isolate A5 and their orthologs on the single-nucleus assemblies of isolates A1 and C2 (Supplementary Table 8). The difference with the previous analysis is that with the single-nucleus data, we are able to identify the putative MAT-type of the A5 haplotypes. We found that for several nucleotype-specific genes (i.e. OG 2995, OG3981 and OG4715), the nuclei from isolates A5 (MAT-3 type) and A1 (MAT-3 type) clustered together (Figure 5c). However, we found that for other nucleotype-specific genes (OG4925, OG4492 and OG4493), the nuclei from isolate A5 (MAT-3 type) clustered with nuclei from isolate C2 which has a MAT-6 type (Figure 5d). The alignments on these nucleotype-specific genes show that A5 nuclei with MAT-3 have similar, but not identical alleles as C2 nuclei (MAT-6 type). In the same way, A5 nuclei with MAT-6 type harbor alleles similar to those of A1 nuclei (MAT-3 type). These results demonstrate that the A5 isolate harbors nucleotypes with regions highly similar to isolates A1 and C2, but that the A1-like alleles are not always found in the same nucleus. The presence of recombinant nucleotypes in isolate A5, involving isolates sharing the same MAT-type, strongly suggest that isolate A5 results from a recombination event between isolates similar to A1 and C2.

Discussion

We identified regions that display drops in coverage in genome assemblies. Within these regions, we identified genes that only displayed a second ortholog in dikaryon isolates and not in monokaryon isolates. We then confirmed with an independent dataset, consisting of a population of individual nuclei, that in dikaryon isolates, genes observed in the regions that displayed a drop in coverage are nucleotype-specific. With the information of nucleotype-specific haplotypes, we found that nucleotypes of isolate A5 were as little 1% diverged from isolate A1 and 4% diverged from isolate C2, suggesting that isolates sharing the same MAT-type as the monokaryon isolates A1 and C2 are closely related ancestors (specially for A1) from which the dikaryon A5 arose. Finally, we identified recombination between nucleotype-specific haplotypes of isolate A5 suggesting that a sexual process involving meiotic-like large scale recombination at some stage is likely at the origin of this isolate.

Our approach allowed us to identify divergent nucleotype-specific idiomorphs in dikaryon isolates situated in different contigs of the short-read whole genome assemblies. This approach differs from previous approaches of global intra-isolate divergence assessment that measured the number of SNPs (Chen *et al.*, 2020) or poly-allelic sites (Wyss *et al.*, 2016). Although the comparison of both types of measurements gave similar information (intra-isolate divergence), their comparison should be carefully addressed as their methodology and the types of sequences compared are different. While SNPs are best identified in low divergence regions, where reads can be confidently mapped to the same contig, these idiomorphic sequences are highly divergent to the point that they are assembled in different contigs in the same genome assembly. Consequently, genetic divergence should be higher in idiomorphs than when the two alleles are collapsed in the genome assembly, resulting in sequences displaying several SNPs.

414 The approach used in this study allowed us to identify nucleotype specific idiomorphs,
415 including the already known HD2 and HD1-like contained in the putative MAT-locus (Ropars
416 *et al.*, 2016). The majority of the nucleotype-specific genes are annotated as hypothetical
417 proteins, with the exception of two galactose oxidases (OG4529, OG4667), a S-adenosyl-L-
418 methionine-dependent methyltransferase (OG1886) and a prephenate dehydrogenase
419 (OG4386). Interestingly, the MAT-locus genes are the only genes identified as assembled
420 into two nucleotype-specific alleles in this study that are common to the two dikaryons. A
421 likely reason why we recovered very few nucleotype-specific genes reflects the choice of
422 filters (syntenic as well as confirmed with single-nucleus data) and the reduction of the
423 complexity of the dataset by keeping only genes with a single copy in the monokaryon
424 isolates A1 and C2. In addition, our methodology did not allow us to identify
425 presence/absence nucleotype-specific haplotypes where a gene is present in one
426 nucleotype and absent in the other nucleotype. Consequently, the count of nucleotype-
427 specific genes represents a lower estimate and we would expect a significantly higher
428 number of actual nucleotype-specific genes between the nuclei in a dikaryon.

429 In AMF, signatures of inter-nucleus recombination in dikaryon isolates have been claimed
430 (Chen *et al.*, 2018a, 2020). Although not specified, the reported large recombination blocks
431 would be consistent with meiotic-type recombination. However, the experimental design on
432 which these results were based did not include any prior crossing of parental isolates or
433 comparison to sequences from putative originating isolates. In the absence of a recent
434 parental mixing, these results could also be the result of processes like mitotic recombination
435 or gene conversion. Furthermore, increased scrutiny revealed technical artifacts suggesting
436 that the method used had limited utility to evaluate inter-nuclear recombination (Auxier &
437 Bazzicalupo, 2019). In this study, we tested for the detection of recombination signatures
438 between homologous regions that could have originated from isolates sharing the same
439 MAT-type. We identified recombinant nucleotypes between different regions of single-
440 nucleus genome assemblies not apparent from the genome assemblies of the dikaryon. The

441 reason for this is that a dikaryon genome assembly lacks phasing information for the
442 haplotypes, contrary to the single-nucleus data where phasing information is immediately
443 apparent (each nucleus has an assigned MAT-type). Integrating the phasing of the
444 haplotypes with the information of the MAT-type, we were able to identify recombination
445 patterns in the single-nucleus data. We did not identify recombination breakpoints, because
446 of the short length of the haplotype sequences. In addition, we did not identify recombination
447 within genes, which is not surprising, as the haplotypes we identified are relatively small, as
448 well as that recombination within genes could disrupt the gene architecture and function.

449 We observed that the A5 MAT-3 nuclei contained three genes more similar to a MAT-6
450 monokaryon, while the A5 MAT-6 nuclei contained three genes more similar to a MAT-3
451 monokaryon. Consequently, the recombination pattern observed is consistent with sexual
452 reproduction because the sequence changes are reciprocal between the two putative
453 parental isolates, although we limited our comparison to only two monokaryon genome
454 assemblies. In contrast, in parasexual recombination, we would expect non-symmetrical
455 recombination patterns leading to loss of heterogeneity between the two different
456 nucleotypes (Forche *et al.*, 2008). This result, coupled with the experimental evidence of
457 molecular activation of genes related to mating in AMF (Mateus *et al.*, 2020), further
458 suggests sexual reproduction in AMF.

459 Our results indicated that if we have different nucleotype-specific haplotypes which are
460 polymorphic in isolate A5 (ie: A/a, B/b) and if there is no recombination, these should be
461 separated into two nucleotypes: MAT-3 (AB) and MAT-6 (ab). However, in contrast, we
462 observed that the dikaryon A5 MAT-3 nucleotype contains a recombined genotype (Ab), and
463 that the A5 MAT-6 nucleotype has a recombined genotype (aB). If we compare the A5 MAT-
464 3 (Ab) and A5 MAT-6 (aB) nucleotypes to A1 MAT-3 (AB) and C2 MAT-6 (ab), the results
465 suggest that A1 and C2-like isolates could be MAT-3 and MAT-6 progenitors of isolate A5,
466 separated by at least one recombinant sexual event. The increased genetic distance (~4%)
467 for C2-like alleles from A5 make it less likely that the actual C2 isolate represent the

progenitor of A5. However, we cannot discount that future sequencing of additional isolates will identify other isolates that could be more closely related to the A5 nucleotypes.

Maintaining monokaryon and dikaryon isolates within the same natural population suggests that both forms are stable over time. Rather than a promiscuous mixing between isolates via anastomosis, a mechanism of recognition that involves the MAT-locus seems to regulate which isolates can form a dikaryon (Corradi & Brachmann, 2017). The fact that nucleotype-specific haplotypes from isolate A5 are more closely related to isolates A1 and C2, and that A5 nucleotypes display recombination, suggests that an isolate similar to A1 and an isolate sharing the same MAT-type as C2 could be the origin of a recombining dikaryon isolate. However, we cannot discard that these findings could apply to another step of the AMF life cycle. It could also be possible that a stable A5 isolate could segregate producing recombined monokaryons highly similar to A1 and that share the same MAT-type as C2, that can disperse and then fuse again to form stable dikaryons and complete a life cycle which involves recombination. It then becomes crucial to experimentally confirm if monokaryon isolates having different MAT-types could generate a dikaryon-like form and if a dikaryon isolate could segregate into recombining monokaryon isolates.

Understanding the life cycle of AMF could have an enormous impact in the generation of AMF genetic variability. The generation of diverse AMF monokaryons or dikaryons could be used to generate variants that enhance plant growth and have an enormous potential in agriculture (Sanders, 2010).

Acknowledgments

We would like to thank Daniel Croll for his feedback on a preliminary version of this manuscript. We also thank three anonymous reviewers that provide constructive feedback

on the manuscript. This study was funded by the Swiss National Science Foundation (Grant number: 31003A_162549 to IRS).

Author contributions

IDM designed and supervised all the analyses. MMSN developed the drop in coverage detection algorithm. JC identified the genes present in the regions displaying drop in coverage. BA developed the genetic relatedness analysis on the nucleotype-specific haplotypes. SL provided valuable comments during all the process. IDM, performed all the bioinformatic analyses, identified the nucleotype-specific genes, performed the phylogenetic analysis, gene-retrieval from genome assemblies and the recombination detection. IDM and IRS wrote the manuscript. All the authors provided valuable contributions and modifications of the manuscript.

Data availability

All data analyzed in this manuscript its available in the NCBI repository and the respective accession identifiers could be found in Supplementary table 1. The custom code developed for the identification of drop in coverage regions could be found in Supplementary file 1.

References

Angelard C, Colard A, Niculita-Hirzel H, Croll D, Sanders IR. **2010**. Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Current Biology* **20**: 1216–21.

Antunes PM, Koch AM, Morton JB, Rillig MC, Klironomos JN. **2011**. Evidence for functional divergence in arbuscular mycorrhizal fungi from contrasting climatic origins. *The New phytologist* **189**: 507–14.

Augé RM. **2001**. Water relations, drought and vesicular-arbuscular mycorrhizal symbiosis. *Mycorrhiza* **11**: 3–42.

Auxier B, Bazzicalupo A. **2019**. Comment on 'Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi'. *eLife* **8**: 1–9.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, *et al.* 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.

Bever JD, Morton J. **1999**. Heritable variation and mechanism of inheritance of spore shape within a population of Scutellosporea pellucida, An arbuscular mycorrhi. *American journal of botany* **86**: 1209–1216.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. **2009**. BLAST+: Architecture and applications. *BMC Bioinformatics* **10**: 1–9.

Chen ECH, Mathieu S, Hoffrichter A, Ropars J, Dreissig S, Fuchs J, Brachmann A, Corradi N. **2020**. More Filtering on SNP Calling Does Not Remove Evidence of Inter-Nucleus Recombination in Dikaryotic Arbuscular Mycorrhizal Fungi. *Frontiers in Plant Science* **11**: 1–9.

Chen EC, Mathieu S, Hoffrichter A, Sedzielewska-Toro K, Peart M, Pelin A, Ndikumana

S, Ropars J, Dreissig S, Fuchs J, *et al.* 2018a. Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. *eLife* **7**: 1–17.

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

Corradi N, Brachmann A. **2017**. Fungal Mating in the Most Widespread Plant Symbionts? *Trends in Plant Science* **22**: 175–183.

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

Emms DM, Kelly S. **2019**. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 1–14.

Forche A, Alby K, Schaefer D, Johnson AD, Berman J, Bennett RJ. **2008**. The Parasexual Cycle in Candida albicans Provides an Alternative Pathway to Meiosis for the Formation of Recombinant Strains (J Heitman, Ed.). *PLoS Biology* **6**: e110.

Giovannetti M, Azzolini D, Citernesi AS. **1999**. Anastomosis Formation and Nuclear and Protoplasmic Exchange in Arbuscular Mycorrhizal Fungi. *Applied and environmental microbiology* **65**: 5571–5575.

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

Harrison MJ. **1997**. The arbuscular mycorrhizal symbiosis: An underground association. *Trends in Plant Science* **2**: 54–60.

van der Heijden M, Klironomos J, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller T,

562 **Wiemken A, Sanders IR**. **1998**. Mycorrhizal fungal diversity determines plant biodiversity,
563 ecosystem variability and productivity. *Nature* **74**: 69–72.

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

566 **Katoh K, Rozewicki J, Yamada KD**. **2017**. MAFFT online service: multiple sequence
567 alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*: 1–7.

568 **Kobayashi Y, Maeda T, Yamaguchi K, Kameoka H, Tanaka S, Ezawa T, Shigenobu S,**
569 **Kawaguchi M**. **2018**. The genome of Rhizophagus clarus HR1 reveals a common genetic
570 basis for auxotrophy among arbuscular mycorrhizal fungi. *BMC Genomics* **19**: 1–11.

571 **Kokkoris V, Chagnon P, Yildirir G, Clarke K, Goh D, MacLean AM, Dettman J, Stefani**
572 **F, Corradi N**. **2021**. Host identity influences nuclear dynamics in arbuscular mycorrhizal
573 fungi. *Current Biology*: 1–8.

574 **Krueger F**. **2015**. Trim galore. *A wrapper tool around Cutadapt and FastQC to consistently*
575 *apply quality and adapter trimming to FastQ files* **516**: 517.

576 **Kumar S, Stecher G, Li M, Knyaz C, Tamura K**. **2018**. MEGA X: Molecular evolutionary
577 genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**: 1547–
578 1549.

579 **Lauri A, Lazzari G, Galli C, Lagutina I, Genzini E, Braga F, Mariani P, Williams JL**. **2013**.
580 Assessment of MDA efficiency for genotyping using cloned embryo biopsies. *Genomics* **101**:
581 24–9.

582 **Leinonen R, Sugawara H, Shumway M**. **2011**. The sequence read archive. *Nucleic Acids*
583 *Research* **39**: 2010–2012.

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

MEM. **00**: 1–3.

Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, Raley C, Miller ME, Silverstein KAT, Henningsen E, *et al.* 2019. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. *Nature communications* **10**: 5068.

Marleau J, Dalpé Y, St-Arnaud M, Hijri M. **2011**. Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. *BMC evolutionary biology* **11**: 51.

Masclaux FG, Wyss T, Mateus-Gonzalez ID, Aletti C, Sanders IR. **2018**. Variation in allele frequencies at the bg112 locus reveals unequal inheritance of nuclei in a dikaryotic isolate of the fungus Rhizophagus irregularis. *Mycorrhiza* **28**: 369–377.

Masclaux FG, Wyss T, Pagni M, Rosikiewicz P, Sanders IR. **2019**. Investigating unexplained genetic variation and its expression in the arbuscular mycorrhizal fungus Rhizophagus irregularis: A comparison of whole genome and RAD sequencing data. *PLoS ONE* **14**: 1–20.

Mateus ID, Rojas EC, Savary R, Dupuis C, Masclaux FG, Aletti C, Sanders IR. **2020**. Coexistence of genetically different Rhizophagus irregularis isolates induces genes involved in a putative fungal mating response. *The ISME Journal*.

Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. **2018**. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**: i142–i150.

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

Morin E, Miyauchi S, San Clemente H, Chen ECH, Pelin A, Providencia I, Ndikumana S,

609 **Beaudet D, Hainaut M, Drula E, *et al.* 2019b**. Comparative genomics of Rhizophagus
610 irregularis , R. cerebriforme , R. diaphanus and Gigaspora rosea highlights specific genetic
611 features in Glomeromycotina. *New Phytologist* **222**: 1584–1598.

612 **Ning L, Liu G, Li G, Hou Y, Tong Y, He J**. **2014**. Current challenges in the bioinformatics of
613 single cell genomics. *Frontiers in Oncology* **4 JAN**: 1–7.

614 **Paradis E, Claude J, Strimmer K**. **2004**. APE: Analyses of phylogenetics and evolution in R
615 language. *Bioinformatics* **20**: 289–290.

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

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

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

624 **Sanders IR**. **2010**. Designer mycorrhizas: Using natural genetic variation in AM fungi to
625 increase plant growth. *ISME Journal* **4**: 1081–1083.

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

630 **Smith SE, Read DJ**. **2008**. *Mycorrhizal Symbiosis*. London: Elsevier Ltd.

631 **Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G,**

Corradi N, Grigoriev I, Gryganskyi A, *et al.* 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**: 1028–1046.

Sullivan MJ, Petty NK, Beatson SA. **2011**. Easyfig: A genome comparison visualizer. *Bioinformatics* **27**: 1009–1010.

Thygesen K, Larsen J, Bødker L. **2004**. Arbuscular mycorrhizal fungi reduce development of pea root-rot caused by Aphanomyces euteiches using oospores as pathogen inoculum. *European Journal of Plant Pathology* **110**: 411–419.

Todd RT, Forche A, Selmecki A. **2017**. Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution. In: The Fungal Kingdom. American Society of Microbiology, 599–618.

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

Yildirir G, Malar C M, Kokkoris V, Corradi N. **2020**. Parasexual and Sexual Reproduction in Arbuscular Mycorrhizal Fungi: Room for Both. *Trends in Microbiology* **xx**: 1–3.

Yilmaz S, Singh AK. **2012**. Single cell genome sequencing. *Current opinion in biotechnology* **23**: 437–43.

Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. **2009**. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Research* **19**: 1586–1592.

Figure legends

Figure 1. Drop in coverage events in isolates A4, A5, A1 and C2. a, Examples of drop in coverage events. We plotted the normalized coverage (y) per position (x). Grey rectangles represent the region detected by the algorithm. The horizontal dashed line represents the normalized coverage. **b,** Number of regions showing a drop in coverage that were detected in each isolate. **c,** Number of genes found in the regions showing a drop in coverage in each isolate. **d,** Summary statistics of regions showing a drop in coverage: i, proportion of total length of regions showing a drop in coverage and proportion of contigs that contain regions with a drop in coverage. ii, Histogram representing the lengths of identified regions where a drop in coverage was detected. iii, Coverage plot on the MAT-locus. Drop in coverage was detected in isolates A4 and A5 but not in A1 and C2.

Figure 2. Identification of orthologs of genes present in regions showing a drop in coverage. a, Orthologous groups that display two or more genes in dikaryons and only a single gene in monokaryons. This analysis was performed independently on isolate A4 and isolate A5. * orthologous genes containing HD2 and HD1-like genes respectively. **b,** Venn diagram representing the number of shared orthologous groups within drop in coverage regions between isolates A4 and A5. Only two orthologous groups were shared between the isolates, they contain the MAT-locus genes HD2 and HD1-like. **c,** Synteny plot between the two contigs containing the different alleles of the MAT-locus of isolates A4 and A5. **d,** Synteny plot between the two contigs containing different alleles of other orthologous genes. Please note that the synteny figures are made from the public available annotations of each

679 genome assembly. Differences in size of open-reading frames (ORF) among isolates are
680 due to differences in detection of ORF on each isolate and likely could be the result of the
681 annotation process.

682

683 **Figure 3. Single-nucleus sequence data confirms that genes contained in regions**
684 **where a drop in coverage was observed are nucleotype-specific.** Phylogenetic
685 reconstruction of single-nucleus for genes found in regions where a drop in coverage was
686 detected in A4 and A5 isolates. The genes are named by their membership to the
687 orthologous groups previously defined. Branch support consisting of 100 bootstraps is
688 shown. When only sequences from three nuclei were included, we performed an UPGMA
689 hierarchical clustering. **a,** data for nuclei from A4 isolate. **b,** data for nuclei from A5 isolate.

690

691 **Figure 4. Nucleotypes from isolate A5 share a more recent evolutionary origin to**
692 **isolates A1 and C2 than among them. a,** Phylogenetic reconstruction of nucleotype-
693 specific alleles in isolate A5 and its orthologs in isolates A1, C2 and A4. **c,** Average genetic
694 distances between the different nucleotype-specific alleles from two alleles from isolates A5
695 and their homologs in isolates A1 and C2. We show histograms representing the genetic
696 distance between the two nucleotype-specific alleles of isolate A5 and their homologues in
697 isolates A1 and C2. For comparisons between A5 and C1 or A1 we used the minimum
698 distance of the two alleles from A5. **d**, Scenarios of genetic similarity between the two A5
699 alleles and alleles from A1 and C2.

700

701 **Figure 5. Recombination events in nucleotype-specific haplotypes in A5 isolate. a,**
702 Schematic representation of possible outcomes after fusion of two different isolates. Please
703 note the schema illustrates different contigs, separated by blank lines and no different

chromosomes. **b,** Multiple sequence alignment of sequences identified in nucleotype-specific haplotypes from whole genome sequence data. We show only a part of the alignment representative of all the aligned sequences. We observe that in all the nucleotype-specific haplotypes, the A5 nucleotypes display only similarity to one of the potential parental isolates. No evidence of recombination is detected within the alignments. **c**, Phylogenetic relationship and multiple sequence alignment on different nucleotype-specific alleles between different nuclei from A1, A5 and C2 isolates. Cases where no recombination was detected. Nuclei having the same MAT-type clustered together. **d**, Phylogenetic relationship and multiple sequence alignment on different nucleotype-specific alleles between different nuclei from A1, A5 and C2 isolates. Cases where recombination was detected. Nuclei having the same MAT-type did not clustered together. We performed 100 bootstraps for the branch support.

Table legends

Table 1. Summary of genes contained in regions displaying a drop in coverage that were validated as nucleotype-specific. We show the number of genes that were found in drop in coverage regions that displayed two copies in the dikaryons and a single copy in the monokaryons. We then filtered out the genes that were present in the same contig. After that we only kept the genes that were present in syntenic regions among the different isolates. Finally, we only kept the genes that displayed at least two types in the single-nucleus genome assemblies. The final number of genes defined as nucleotype-specific was validated during the previous sequential steps.

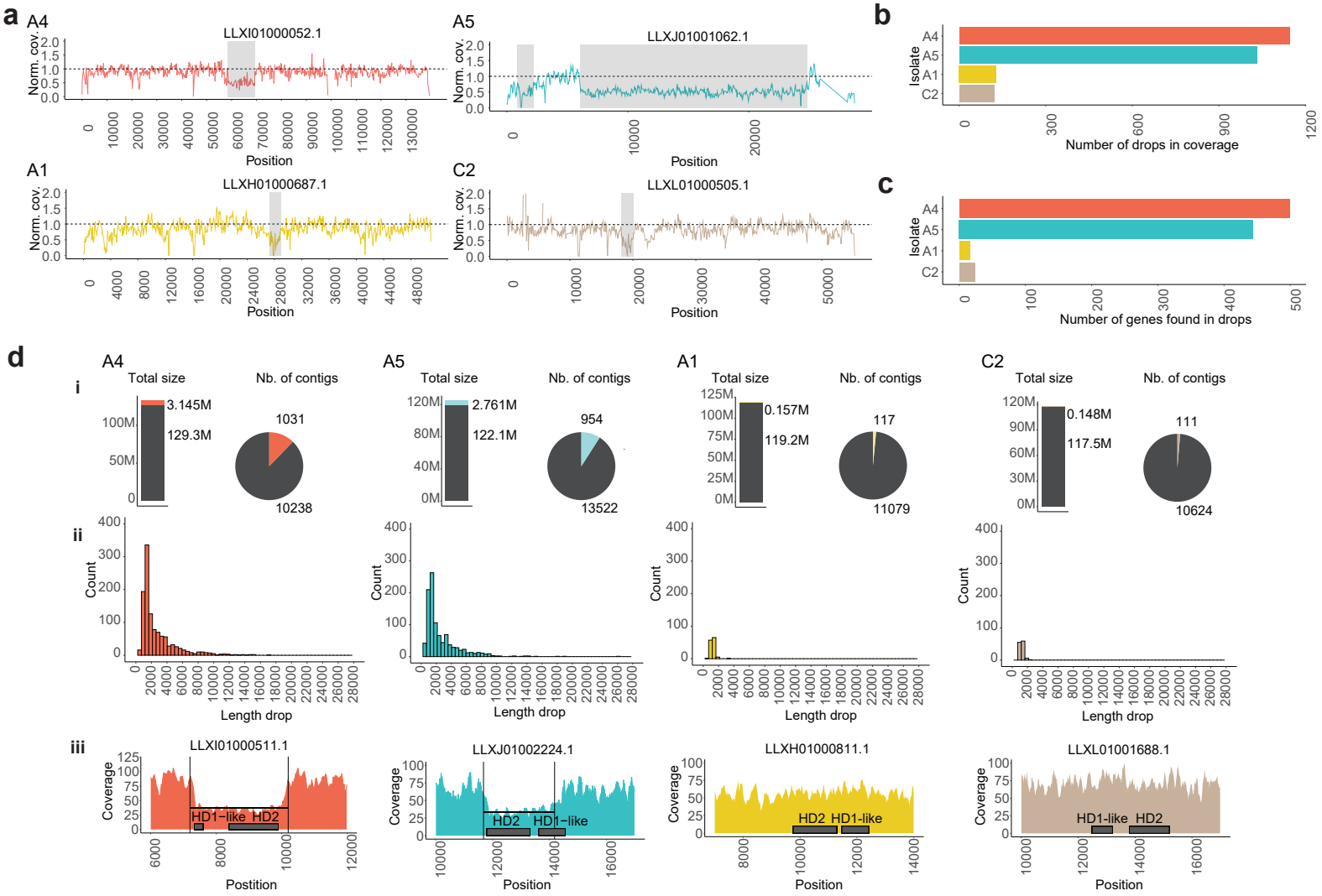
729

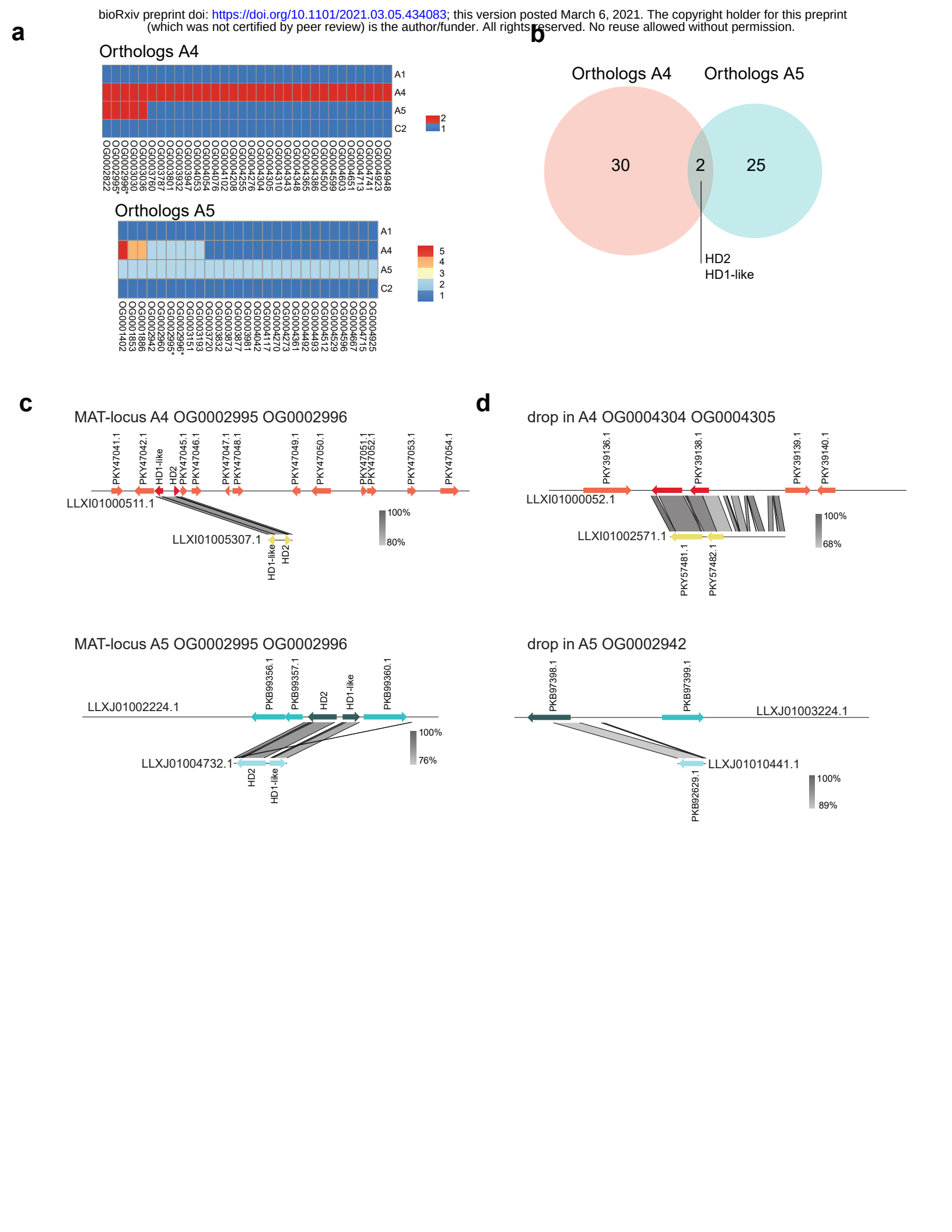
730

731

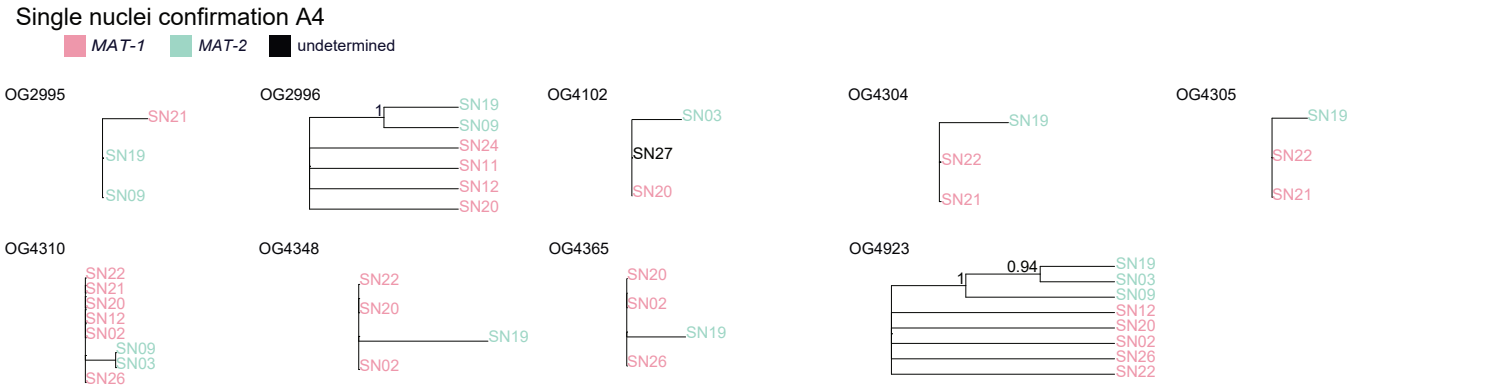
732

733

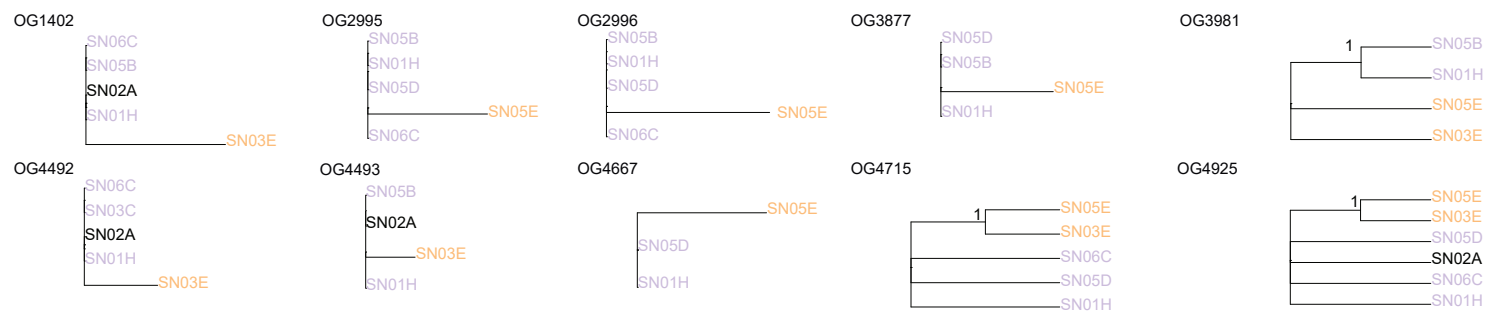

a
A4
LLXI01000052.1
A5
LLXJ01001062.1
A1
LLXH01000687.1
C2
LLXL01000505.1
Norm. cov.
Position
b
Isolate
Number of drops in coverage
c
Number of genes found in drops
d
Total size
Nb. of contigs
3.145M
129.3M
1031
10238
2.761M
122.1M
954
13522
0.157M
119.2M
117
11079
0.148M
117.5M
111
10624
Count
Length drop
Coverage
Postition
HD1−like HD2
HD2 HD1−like
HD2 HD1-like
HD1-like HD2



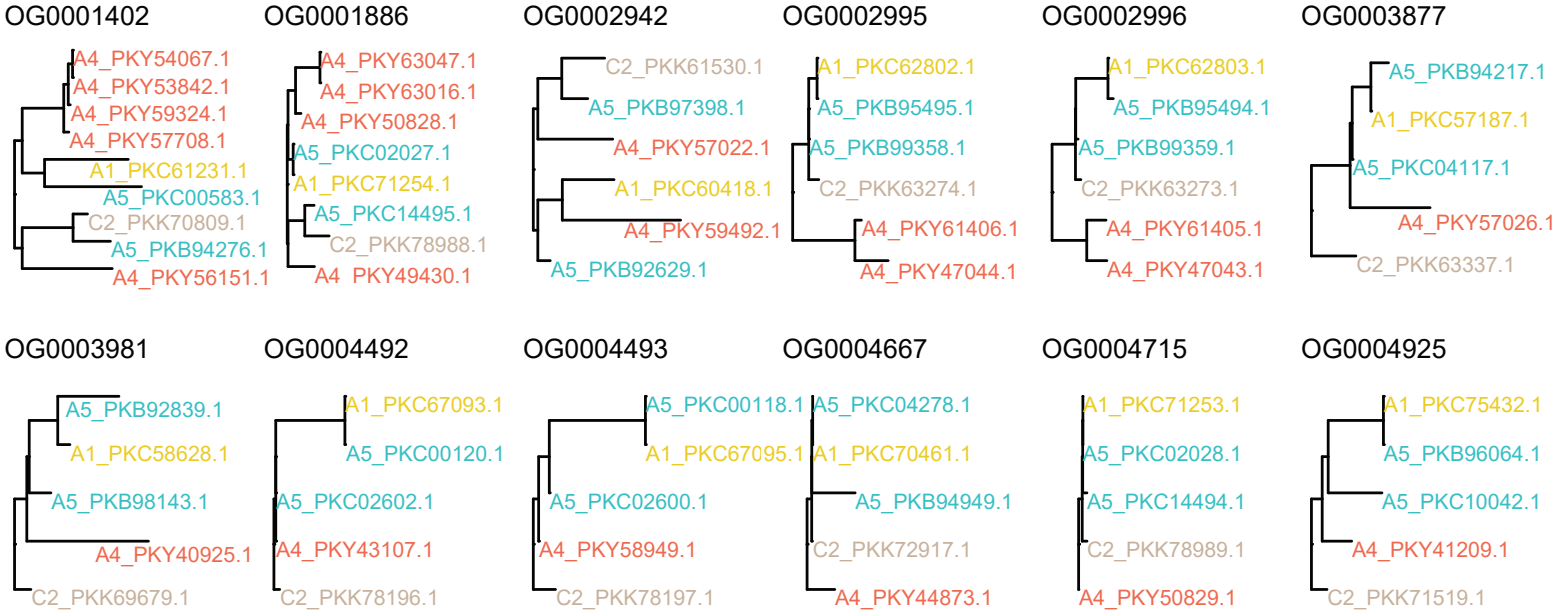
a



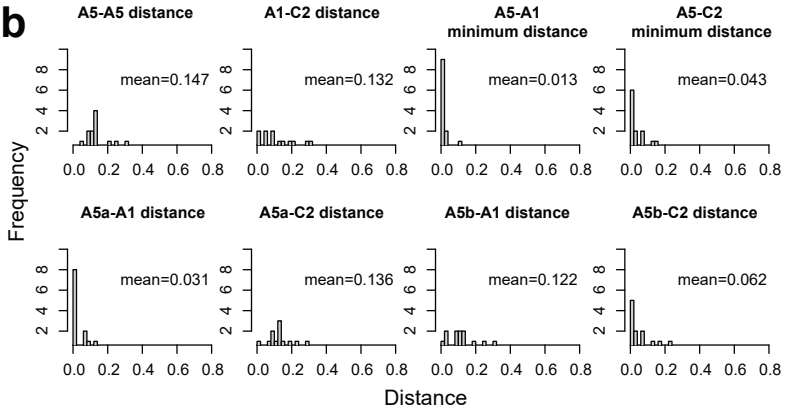
b



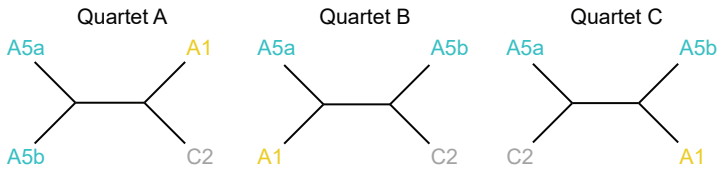
a



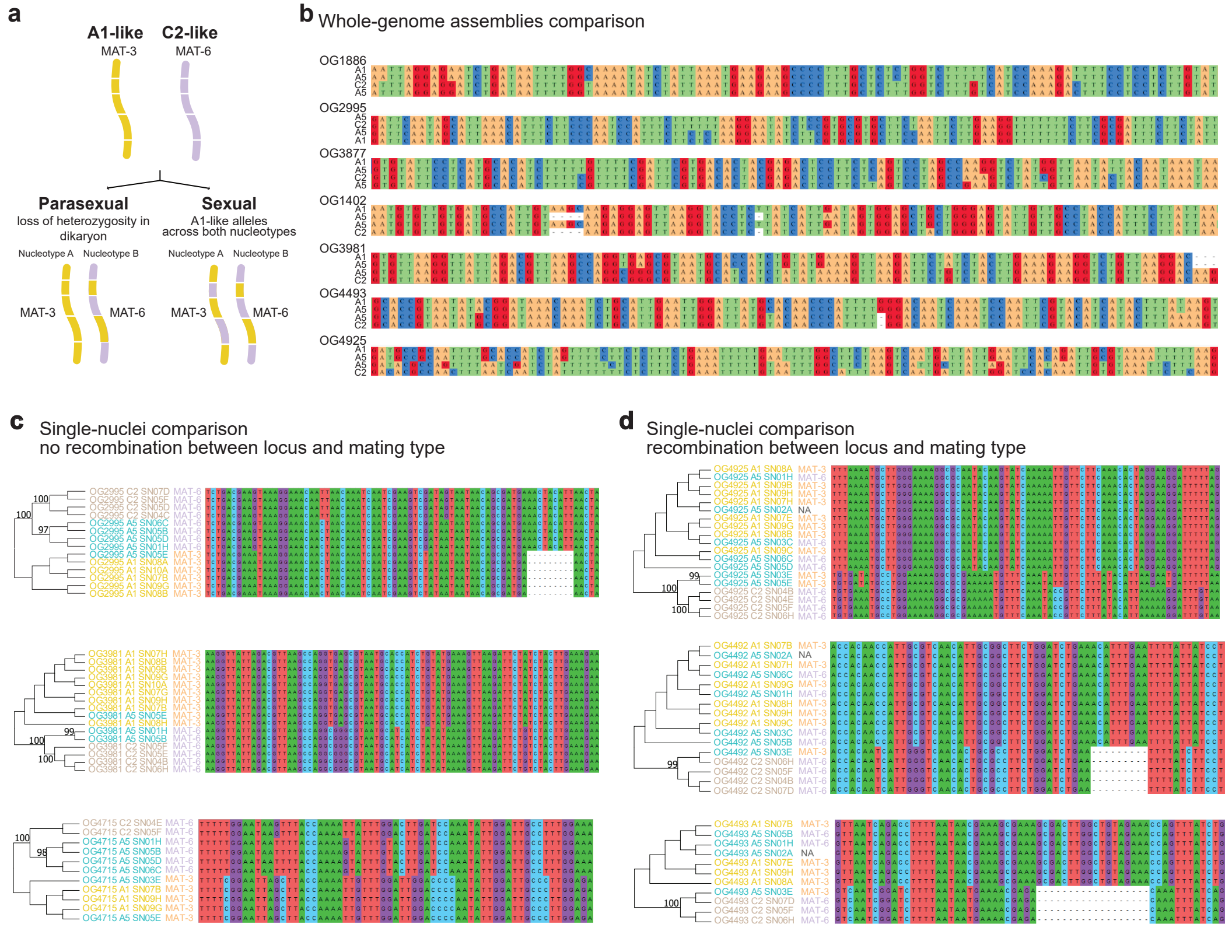
b



c



Scenario	Quartet	Count
A5 alleles more closely related	A	0
A5 alleles more distantly related	B or C	12


a
A1-like
MAT-3
C2-like
MAT-6
Parasexual
loss of heterozygosity in dikaryon
Sexual
A1-like alleles across both nucleotypes
Nucleotype A
Nucleotype B
MAT-3
MAT-6
b
Whole-genome assemblies comparison
OG1886
OG2995
OG3877
OG1402
OG3981
OG4493
OG4925
c
Single-nuclei comparison
no recombination between locus and mating type
OG2995 C2 SN07D MAT-6
OG2995 C2 SN05F MAT-6
OG2995 C2 SN05D MAT-6
OG2995 C2 SN04C MAT-6
OG2995 A5 SN06C MAT-6
OG2995 A5 SN05B MAT-6
OG2995 A5 SN05D MAT-6
OG2995 A5 SN01H MAT-6
OG2995 A5 SN05E MAT-3
OG2995 A1 SN08A MAT-3
OG2995 A1 SN10A MAT-3
OG2995 A1 SN07B MAT-3
OG2995 A1 SN09G MAT-3
OG2995 A1 SN08B MAT-3
100
100
97
OG3981 A1 SN07H MAT-3
OG3981 A1 SN08B MAT-3
OG3981 A1 SN09B MAT-3
OG3981 A1 SN09G MAT-3
OG3981 A1 SN10A MAT-3
OG3981 A1 SN07G MAT-3
OG3981 A1 SN09H MAT-3
OG3981 A1 SN07B MAT-3
OG3981 A5 SN05E MAT-3
OG3981 A1 SN08H MAT-3
OG3981 A5 SN01H MAT-6
OG3981 A5 SN05B MAT-6
OG3981 C2 SN05F MAT-6
OG3981 C2 SN05E MAT-6
OG3981 C2 SN04B MAT-6
OG3981 C2 SN06H MAT-6
99
100
100
OG4715 C2 SN04E MAT-6
OG4715 C2 SN05F MAT-6
OG4715 A5 SN01H MAT-6
OG4715 A5 SN05B MAT-6
OG4715 A5 SN05D MAT-6
OG4715 A5 SN06C MAT-6
OG4715 A5 SN03E MAT-3
OG4715 A1 SN07B MAT-3
OG4715 A1 SN09H MAT-3
OG4715 A1 SN09G MAT-3
OG4715 A5 SN05E MAT-3
100
98
d
Single-nuclei comparison
recombination between locus and mating type
OG4925 A1 SN08A MAT-3
OG4925 A5 SN01H MAT-6
OG4925 A1 SN09B MAT-3
OG4925 A1 SN09H MAT-3
OG4925 A1 SN07H MAT-3
OG4925 A5 SN02A NA
OG4925 A1 SN07E MAT-3
OG4925 A1 SN09G MAT-3
OG4925 A1 SN08B MAT-3
OG4925 A5 SN03C MAT-6
OG4925 A1 SN09C MAT-3
OG4925 A5 SN06C MAT-6
OG4925 A5 SN05D MAT-6
OG4925 A5 SN03E MAT-3
OG4925 C2 SN04B MAT-6
OG4925 C2 SN04E MAT-6
OG4925 C2 SN05F MAT-6
OG4925 C2 SN06H MAT-6
99
100
100
OG4492 A1 SN07B MAT-3
OG4492 A5 SN02A NA
OG4492 A1 SN07H MAT-3
OG4492 A5 SN06C MAT-6
OG4492 A1 SN09G MAT-3
OG4492 A5 SN01H MAT-6
OG4492 A1 SN08H MAT-3
OG4492 A1 SN09H MAT-3
OG4492 A1 SN09C MAT-3
OG4492 A5 SN03C MAT-6
OG4492 A5 SN05B MAT-6
OG4492 A5 SN03E MAT-3
OG4492 C2 SN06H MAT-6
OG4492 C2 SN05F MAT-6
OG4492 C2 SN04B MAT-6
OG4492 C2 SN07D MAT-6
99
OG4493 A1 SN07B MAT-3
OG4493 A5 SN05B MAT-6
OG4493 A5 SN01H MAT-6
OG4493 A5 SN02A NA
OG4493 A1 SN07E MAT-3
OG4493 A1 SN09H MAT-3
OG4493 A1 SN08A MAT-3
OG4493 A5 SN03E MAT-3
OG4493 C2 SN07D MAT-6
OG4493 C2 SN05F MAT-6
OG4493 C2 SN06H MAT-6
100