Mycorrhiza (2018) 28:369–377
https://doi.org/10.1007/s00572-018-0834-z

ORIGINAL ARTICLE

Variation in allele frequencies at the *bg112* locus reveals unequal inheritance of nuclei in a dikaryotic isolate of the fungus *Rhizophagus irregularis*

Frédéric G. Masclaux[1,2] · Tania Wyss[1] · Ivan D. Mateus-Gonzalez[1] · Consolée Aletti[1] · Ian R. Sanders[1]

Received: 16 January 2018 / Accepted: 5 April 2018 / Published online: 19 April 2018

Abstract

The genetic state of the arbuscular mycorrhizal fungus species *Rhizophagus irregularis* differs among isolates, including both homokaryotic and dikaryotic isolates. Via the production of multi-nucleate axexual spores, siblings of dikaryotic isolates may inherit unequal frequencies of nucleotypes. Using *bg112*, a microsatellite marker, previous studies revealed that lines deriving from single spores of the dikaryotic *R. irregularis* isolate C3 differed in their proportions of different alleles. A genomic study of single nuclei of *R. irregularis*, however, suggested that this marker was a multi-copy locus and that therefore it was inappropriate to study the inheritance of nuclei in dikaryotic isolates. In this study, we first analysed whole genome data of several *R. irregularis* isolates and demonstrated that *bg112* is indeed a single copy locus in these genomes. Thus, the *bg112* locus is a suitable marker to study the relative frequency of nucleotypes in *R. irregularis*. Second, by using amplicon sequencing, we confirmed the existence of one allele of *bg112* in two homokaryotic isolates (DAOM197198 and C2) and two alleles in the dikaryotic isolate (C3). Finally, we found that the relative proportions of two *bg112* alleles differed significantly among dikaryotic single-spore lines derived from isolate C3, indicating that genetically different nucleotypes are inherited unequally in this dikaryotic *R. irregularis* isolate.

Keywords Arbuscular mycorrhizal fungi · Single-spore lines · *Rhizophagus irregularis* · Nucleotype frequency Nucleus inheritance

Introduction

Arbuscular mycorrhizal fungi (AMF) form symbioses with the roots of around 80% of terrestrial plants, influencing plant performance (Smith and Read 2008) and plant community diversity (van der Heijden et al. 1998). These fungi are coenocytic, and a single-nucleus stage has never been observed in the life cycle of these fungi (Sanders and Croll 2010). Recent genome sequencing studies have shed light on the genomic organisation of these fungi. Whole genome sequencing of *Rhizophagus irregularis* isolates recently showed that some isolates are homokaryotic, thus, harbouring genetically identical nuclei (Tisserant et al. 2013; Lin et al. 2014; Ropars et al. 2016). Two other isolates independently isolated from the same geographic area as the homokaryotic isolates appear to be dikaryotic, i.e. they harbour a population of two genetically different nuclei (nucleotypes), present in equal proportions (Ropars et al. 2016). In dikaryotic isolates, the mechanisms of nuclei inheritance in the spore are currently unknown. The two genetically distinct nuclei either could be inherited in equal proportions in each newly formed spore, or unequal proportions of the two nucleotypes could be inherited in sibling spores.

Nucleus inheritance was previously studied in *R. irregularis* isolate C3. This isolate was recently shown to be a clone of the dikaryotic isolate A4, as no polymorphism between C3 and A4 was detected at 13,184 sites, distributed across the genome (Wyss et al. 2016; Ropars et al. 2016). The two genetically

Frédéric G. Masclaux and Tania Wyss contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00572-018-0834-z) contains supplementary material, which is available to authorized users.

✉ Ian R. Sanders
ian.sanders@unil.ch

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

[2] Vital-IT Group, Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

different nuclei that co-exist in C3 could be inherited in unequal proportions by random sorting during spore formation. Single-spore lines (SSLs), each generated by single-spore propagation of isolate C3, were shown to differ strongly in their quantitative growth traits, as well as in their effects on rice growth (Angelard et al. 2010; Ehinger et al. 2012). Further investigations revealed that the SSLs differed in their relative frequency of alleles at the *bg112* locus (Angelard et al. 2010; Ehinger et al. 2012). This was interpreted as likely representing a difference in the frequency of the distinct nucleus genotypes among the SSLs.

The generation of such interesting quantitative trait variation among SSLs of *R. irregularis* isolate C3 is intriguing because these fungi are thought to produce spores clonally by asexual reproduction and likely did so in the experiments performed by Angelard et al. (2010). It is hypothesised that *R. irregularis* also may be capable of reproducing sexually (Croll and Sanders 2009; Riley and Corradi 2013; Ropars et al. 2016). Nevertheless, in the experiments of Angelard et al. (2010) and Ehinger et al. (2012), it is unlikely that SSLs were the result of any sexual reproduction during the course of the experiments in in vitro conditions. Additionally, large-scale sequencing of three biological replicates of each of 19 in vitro cultured *R. irregularis* from a Swiss population (including C3) revealed that independently cultured replicates of the same isolate were genetically identical, indicating no recombination during culturing (Wyss et al. 2016). Furthermore, clonal reproduction of *R. irregularis* likely is very common in nature because isolates sharing the same allelic composition have been found in different parts of the same field from which those isolates originated (Croll et al. 2008). A recent study of *R. irregularis* isolates using several thousand single-nucleotide polymorphisms found identical clones of *R. irregularis* at an inter-continental scale, including some of those from the same Swiss population, indicating that clonal reproduction in these fungi seems to be a widespread phenomenon (Savary et al. 2018).

R. irregularis commonly is propagated in vitro by transferring thousands of spores and hyphae from a mature culture to a new Petri dish (Rosikiewicz et al. 2017). Transferring thousands of spores and hyphae at once likely does not alter the nucleus frequencies in the resulting mycelium, because all nucleotypes are transferred at once from one generation to the next. In contrast, generation of single-spore lines via single-spore propagation may constitute a bottleneck, and by stochastic processes, variation in nucleus frequencies might arise among SSLs (Angelard et al. 2010; Ehinger et al. 2012).

The estimation of the proportions of genetically different nuclei among SSLs requires a sensitive and accurate approach. While sequencing of single nuclei is appropriate for analysing whether an AMF isolate is homokaryotic or dikaryotic, it is not practical for addressing the question of whether the frequency of different nucleus genotypes differs among SSLs or not. First, it is not easy to produce enough DNA libraries, each of a high enough quality for whole genome sequencing, from individual nuclei. For example, Lin et al. (2014) were able to successfully amplify DNA and make 40 libraries from isolated nuclei. Subsequent quality control of the libraries revealed that only four single nucleus libraries were suitable for sequencing. Second, to measure frequencies of two genetically different nucleus genotypes among SSLs would require obtaining a high-quality single-nucleus whole genome assembly from each of a very large number of nuclei from each of the SSLs. Finally, in order to measure the relative proportions of nucleus genotype-specific alleles, a high depth of coverage would be required for the whole genome of each single nucleus. The preparation and high depth sequencing of such a large number of libraries would be extremely expensive.

An effective approach to quantify the proportions of genetically different nuclei in an isolate is to target a single-copy locus that has nucleus-specific alleles. The nucleus-specific allele approach was used previously at a microsatellite locus called *bg112*. Relative differences in *bg112* allele frequencies at this locus were observed among single-spore lines of *R. irregularis* (Angelard et al. 2010; Ehinger et al. 2012). The differences in *bg112* allele frequencies likely reflected changes in the relative proportions of nuclei carrying different alleles at the *bg112* locus. Recently, genome sequencing of four individual nuclei of *R. irregularis*, isolate DAOM197198, revealed that three so-called *bg112* loci, with divergent sequences, were present in a single nucleus, with no copy number variation among nuclei (Lin et al. 2014). This result raised doubts about the suitability of the *bg112* locus to study the variation in nucleus frequencies. Previous studies on *bg112* allele frequency variation were conducted on C3, however, and not on DAOM197198, where there may be a different number of copies. It previously was shown that the copy number of genes other than *bg112* differs among isolates of *R. irregularis* (Corradi et al. 2007).

Because of the doubts about the suitability of the *bg112* marker system, we first verified its suitability to study nucleus frequency changes in *R. irregularis*. Specifically, we used in silico and molecular methods to demonstrate that the *bg112* marker previously studied by Angelard et al. (2010) represents a single-copy locus in the published genomes of several *R. irregularis* isolates. We then addressed how many alleles at the *bg112* locus can be found in three different *R. irregularis* isolates (C2, C3 and DAOM197198) and whether or not their relative frequencies differed among samples of SSLs deriving from the dikaryotic isolate C3. We chose C3 because it had been used for previous studies on the inheritance of nuclei and we chose isolates C2 and DAOM 197198 because they both were reported to be homokaryotic and because whole genome sequence data existed for both isolates.

Materials and methods

Genome assemblies and evaluation of their quality

The genome of *R. irregularis*, being AT-rich and containing many repeated regions, is difficult to assemble. Nevertheless, in order to study how many copies of *bg112* were present in the genomes, we first had to ensure that the available genome assemblies were of good quality. We investigated 14 published genome assemblies in this study. The origin of sequencing reads and assemblies of *R. irregularis* is described in Supplementary Table S1. In addition to these published assemblies, we used the initial cleaned sequencing reads originating from four of the assemblies (DNA1, DNA2, N6 and N36) in Lin et al. (2014) to make new assemblies using the SPAdes genome assembler version 3.6.2 with the 'careful' mode (Bankevich et al. 2012). The objectives of making new assemblies were to use a recent and improved genome assembler and to obtain assemblies made with the same method as other published genomes (Supplementary Table S1). Quality characteristics of each assembly were assessed with Quast 4.2 (Gurevich et al. 2013). In addition, the quality of the assemblies was evaluated by mapping the set of 579,135 expressed sequence tags (ESTs) described in Tisserant et al. (2012) with BLAT v3.4. First, the proportion of transcripts that map to an assembly should reflect the level of completeness of the assembly. Second, the number of matches per transcript should not differ strongly across assemblies.

Identification of bg112 *loci in genome assemblies*

We used eight *bg112* sequences that were amplified and sequenced in the study by Angelard et al. (2010) as queries for a BLAST search in all available assemblies. The accession numbers of the *bg112* sequences in GenBank are GU930826, GU930824, GU930827, GU930828, GU930825, GU930836, GU930835, GU930834. We also used the Bg112-d and Bg112-up primer sequences in the query to identify the potential *bg112* loci that could actually be amplified by these primers. A word size of 7 and a threshold e-value of 10 were set up for all BLASTs. The hits were subsequently filtered to retain only the hits longer than 15 bp with at least 90% identity for the primers and the hits longer than 80 bp with at least 80% identity for the *bg112* sequences.

The number of targets that theoretically can be amplified with the Bg112-d and Bg112-up pair of primers was evaluated with both in silico PCR (Kuhn et al. 2013) and custom Perl scripts.

Biological material and DNA extraction

We extracted the DNA from the *Rhizophagus irregularis* reference isolate DAOM197198 and from 2 isolates, C2 and C3, that originated from a field in Switzerland (Koch et al. 2004). C3 is considered to be genetically identical to A4 (Wyss et al. 2016). We also extracted DNA from eight single-spore lines derived from isolate C3 (C3-SSL, C3-SSL-d, C3-SSL-e, C3-SSL-f, C3-SSL-2, C3-SSL-2a, C3-SSL-2d and C3-SSL-2e). Methods to generate and maintain SSLs were described elsewhere (Angelard et al. 2010). A graphical view of the origin of, and relationships among, the different lines is shown in Supplementary Fig. S1. All *R. irregularis* isolates and SSLs were grown in vitro in association with Ri T-DNA transformed carrot roots (Bécard and Fortin 1988). Roots and isolates were co-cultured in split-plates with a filter (Durapore membrane filter, 0.45 μm HV, Merck Millipore Ltd.) covering the hyphal compartment to allow the collection of fungal material free of carrot DNA (Rosikiewicz et al. 2017). After 4 months of growth at 25 °C, the medium from the fungal compartments of the split-plates, containing mycelium and spores, was dissolved in citrate buffer (450 ml ddH_2O, 8.5 ml 0.1 N citric acid, 41.5 ml 0.1 N Na-citrate) for 1.5 h and washed with sterile double-deionised water (ddH_2O). Each DNA sample was extracted from material from one fungal compartment. Each mycelium sample was ground in liquid nitrogen with a pellet pestle (Fisher Scientific). The DNeasy Plant Mini kit (Qiagen) was used for DNA extraction following the manufacturer's instructions, and the DNA was recovered in 60-μl AE buffer.

Identification of different bg112 *alleles by amplicon sequencing*

We used amplicon sequencing to determine the number of different *bg112* alleles in isolates DAOM197198, C2 and C3. The *bg112* locus was amplified from DNA of these isolates using the primers Bg112-up and Bg112-d. We included a six nucleotide-long barcode at the 5′-end of the primer Bg112-up so that we could pool all PCR products into one single DNA library for sequencing. The PCR reaction was set up as follows: 2 ng of genomic DNA, 2.5 μl 5 × Q5 HF DNA polymerase buffer, 0.1 μl 25 mM dNTP, 2.5 μl 5 × high GC enhancer, 2.5 μl 1 μM primer Bg112-up, 2.5 μl 1 μM primer Bg112-d, 0.2 μl Q5 HF DNA polymerase (2000 U/ml, NEB) and 1.2 μl ddH2O. Cycling conditions were as follows: 4 min at 98 °C, 30 cycles of 30 s at 98 °C, 1 min at 60 °C, 1 min at 72 °C and a final step for 2 min at 72 °C. The PCR conditions were stringent to ensure high specificity of the primers. Amplification success was verified by running 3 μl of each PCR reaction on a 1% agarose gel in 1 × TBE, stained with ethidium bromide, for 1 h at 100 V. PCR products were then pooled into one tube. We concentrated the pooled DNA by ethanol precipitation and purified it with a 1.5 × ratio of Agencourt AMPure XP magnetic beads followed by elution with 50 μl ddH_2O. The DNA was then converted into a sequencing-compatible library following the TruSeq DNA library preparation protocol (Illumina) without DNA

fragmentation. A second amplicon sequencing library was prepared using the same protocol and barcoded primers as described above but using independent genomic DNA samples. The two amplicon sequencing libraries were sequenced on two different lanes of an Illumina MiSeq sequencer (2 × 250 bp paired-end reads) at the Lausanne Genomic Technologies Facility, Switzerland.

Analysis of amplicon sequencing data

Sequences from the two libraries were analysed independently. We processed the raw reads with the script Tagcleaner.pl to trim Illumina adapters (Schmieder et al. 2010). Reads were quality-filtered with the script PrinSeqlite.pl version 0.20.4 (Schmieder and Edwards 2011) to trim low-quality 3′-ends and to remove reads with low quality or containing uncalled bases (N). Reads were then de-multiplexed according to barcode sequences with a custom Perl script. Because amplicon length is shorter than read length, paired-end reads were split into single-end reads. Single-end reads 1 and single-end reads 2 were processed independently. Amplicon sequences were kept only when both forward and reverse primers could be identified fully or partially at the extremities of each sequence. We then counted and recorded unique sequences for each sample and calculated the occurrence of each sequence in the whole dataset. Only the 50 most frequent sequences in the whole dataset were retained for the following steps. In each sample, an allele was considered as real if its frequency was more than 4% of the total number of sequences in the sample. The contribution of each sample to the total amount of each allelic sequence observed in the library was calculated. In addition, only alleles representing at least 2.5% of the total sequence counts were considered. This very conservative filtering was applied to distinguish reliable alleles from artefactual sequences and also to avoid any risk of leakage or contamination. We reported the occurrence of all recorded alleles per locus. This included the ones that were probably present because of an artefactual stutter effect.

Allele frequency measurements at the bg112 *locus in single-spore lines*

The *bg112* locus was amplified in *R. irregularis* isolate C3 and eight C3-derived SSLs using the primers Bg112-up and Bg112-d that were previously reported by Angelard et al. (2010). The DNA was extracted from one mycelium sample of each SSL following the methods described above. PCR conditions were the same as described for amplicon sequencing, except that the Bg112-up primer did not include a barcode sequence and was tagged with a fluorophore (FAM), and that a different *Taq* polymerase was used (GoTaq, Promega). The *bg112* locus was amplified four times separately for each DNA sample, with 2 ng of genomic DNA in each PCR, to yield four technical replicates per SSL DNA sample. For each PCR product, alleles were separated based on fragment length by capillary electrophoresis on a capillary automated sequencer (3130xl Genetic Analyser, Applied Biosystems). Alleles were identified and sized with the GeneMapper software (Applied Biosystems) and the heights of peaks, measured as relative fluorescence units (RFU), were recorded for each allele. This method was previously used by Angelard et al. (2010) and Ehinger et al. (2012), where it was tested rigorously for accuracy in quantifying known proportions of different amplicons and also shown to be reliable with little variation in peak heights among replicate samples. The same approach has been used in many other studies of allele frequencies, especially in pooled samples. See supplementary information in Angelard et al. (2010) for information about tests of reliability and the relevant references to other studies. Artefact stutter peaks are common in microsatellite marker analysis and can appear in the *bg112* sequence mainly at two locations. First, at the "TTTTT" region between the primer Bg112-d and the microsatellite, resulting in a DNA fragment that is shorter by one nucleotide (i.e. missing one T). Second, at the microsatellite region, resulting in DNA fragments that are longer or shorter by three nucleotides (an additional or missing TTA). We applied a quality filter to each capillary electrophoresis result: the ratio between the height of the peak at 208 bp (allele 1 in isolate C3) and the height of the peak at 207 bp (allele 1 with stutter effect) had to be greater than 3 to consider the result as valid. In samples that harboured more than one allele, we used the peak height data to calculate the proportion of a given allele relative to the other. The reliability of this method was described previously (Angelard et al. 2010; Ehinger et al. 2012).

We tested whether there were significant differences in the proportion of allele 1 among the DNA samples of the C3 isolate and the eight SSLs. The proportions were transformed to arcsin values because proportions have fixed bounds (0 and 1). We applied a general linear model ANOVA to determine if the proportion of allele 1 differed among samples of SSLs, using the four technical PCR replicates of each SSL as replicates. Post hoc pair-wise comparisons between SSLs were conducted using the Tukey's HSD test implemented in the *glht* function from the 'multcomp' package for R (Hothorn et al. 2016). Details of the R code are reported in Supplementary Results.

Results

Quality of genome assemblies

We evaluated the quality of 20 genome assemblies of the *R. irregularis* isolates DAOM197198, A1, A4, A5, B3 and C2 (Supplementary Table S1). The assessment of the genome assembly quality is important for obtaining an accurate measurement of the copy number of the *bg112* and *bg112-like* loci.

Very incomplete or fragmented assemblies could be missing additional *bg112* loci. Additional assemblies were generated from DAOM197198 data with SPAdes, which is among the most efficient assemblers, in particular for microsatellite regions. Most of the assemblies appeared to be quite complete despite being very fragmented. For example, more than 90% of ESTs reported by Tisserant et al. (2012) mapped to all assemblies, except single-nucleus assemblies N31, N33 and N36, which indicated that most assemblies did not differ greatly in their quality (Supplementary Table S1 and Supplementary Fig. S2). Re-assembling genomes with SPAdes did not increase their completeness compared to using the SOAPdenovo2 genome assembler (Supplementary Fig. S2).

The **bg112** *marker represents a single locus in several genome assemblies of R. irregularis*

The *bg112* marker comprises a tri-nucleotide repeat (Fig. 1a), with alleles that differ in the number of repeats and other deletions, insertions and substitutions. Previous studies of allele frequencies at the *bg112* locus were based on DNA amplified using the primer pair Bg112-up and Bg112-d (Angelard et al. 2010). Following Angelard et al. (2010), we refer to the *bg112* marker as the sequences that can be amplified with this primer pair. Lin et al. (2014) observed the existence of several *bg112-like* sequences on different scaffolds of the DAOM197198 genome by conducting a BLAST search using one of the *bg112* alleles

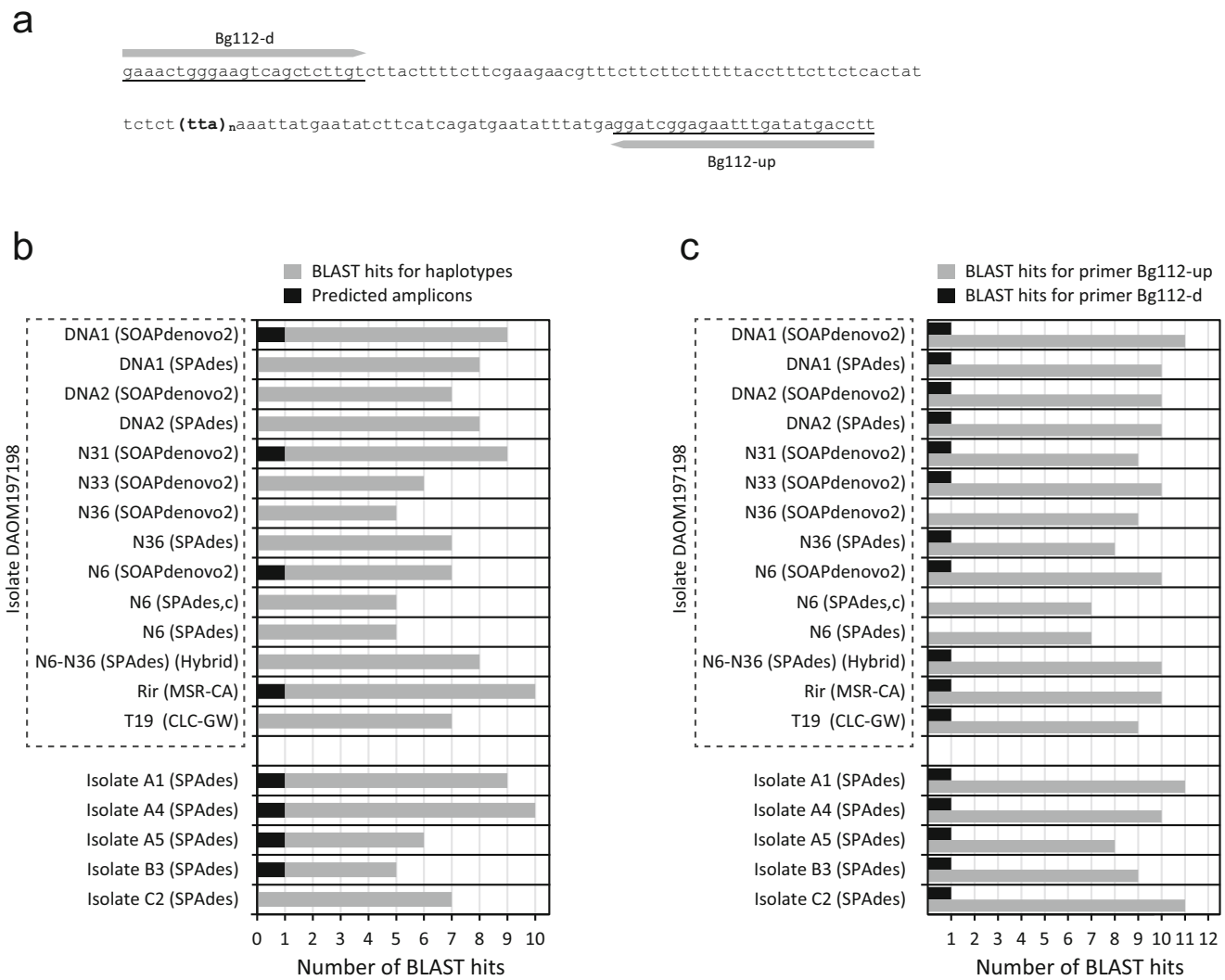

Fig. 1 Analysis of *bg112-like* sequences and the *bg112* marker in *R. irregularis* assemblies. **a** The typical *bg112* marker sequence. The positions of primers Bg112-up and Bg112-d are underlined and indicated by grey arrows. The repetitive microsatellite region is represented with the standard motif between brackets with *n* ranging between 24 and 40 trinucleotide motifs. **b** Number of BLAST hits of the *bg112*-like sequence (grey bars) and predicted number of in silico-predicted amplicons using primers Bg112-up and Bg112-d among the number of BLAST hits (black bars). The reported numbers of predicted amplicons are based on two different approaches, i.e. with an in silico PCR program (Kuhn et al. 2013) and with custom Perl scripts, which provided identical results. **c** Numbers of BLAST hits for each primer sequence in a given assembly. The bars in **b** and **c** indicate the number of occurrences

documented by Angelard et al. (2010). Indeed, performing a BLAST with a set of eight *bg112* sequences led us to identify between five and ten copies of sequences similar to *bg112*, depending on the assembly and the fungal isolate (Fig. 1b). We confirmed the existence of several loci, with some sequence similarity to bg112, in DAOM197198, as reported by Lin et al. (2014). High sequence divergence, however, was found among these sequences (Supplementary Figs. S3 and S4). The Bg112-up primer sequence was detected from seven to eleven times in the different assemblies and the different fungal isolates. In contrast, the Bg112-d primer sequence was never found more than once (Fig. 1c). No match of this primer was found in the N36 assembly made with SOAPdenovo2 and two N6 assemblies, possibly because these assemblies are less complete than some of the others. Interestingly, the N36 assembly made with SPAdes revealed a match of the primer Bg112-d although no match was found with the N36 assembly made with SOAPdenovo2. The match of Bg112-d was perfect in all assemblies where there was a BLAST hit but not in the C2 assembly. It is likely that the typical locus *bg112* was not assembled in C2. In silico PCR revealed that the primer Bg112-d ensures that PCR is specific to a single *bg112* locus in these genomes (Fig. 1b). Therefore, only a single *bg112* PCR target was found in eight assemblies, including genome assemblies of other isolates than DAOM197198 (Fig. 1b). Repeated microsatellite regions, such as $(tta)_n$, are problematic for genome assembly software that frequently leaves a gap in these regions. Analysis of the assemblies showed that a single *bg112* locus is targeted in all available *R. irregularis* genomes when using the *bg112* marker system reported in previous studies by Angelard et al. (2010) and Ehinger et al. (2012).

R. irregularis isolates harbour different numbers of alleles at the single **bg112** locus

To identify and characterise different alleles at the specific *bg112* locus, we performed ultra-deep sequencing of PCR products (amplicon sequencing) amplified with the Bg112-up and Bg112-d primers in DNA from three *R. irregularis* isolates (DAOM197198, C2 and C3). The analysis was based on sequence length variation using very stringent filters. One predominant sequence was found in DAOM197198 and C2, indicating that a single allele exists at the *bg112* locus in these two isolates (Fig. 2; Supplementary Fig. S5). In C3, we detected two predominant sequences that differed in length, indicating that two alleles exist at the *bg112* locus in this isolate (Fig. 2; Supplementary Fig. S5). Allele length differed among all isolates, with DAOM197198 harbouring the longest allele, C3 harbouring two alleles of intermediate length and C2 harbouring the shortest allele. Additional potential alleles that differed from the dominant alleles were detected in C2 and C3, but they did not pass the strict filters (see "Materials and methods" for details). The sequence of all dominant alleles corresponded to the expected sequences from their respective genome assemblies (Supplementary Fig. S6).

Proportion of **bg112** *alleles differs among samples of single-spore lines*

Angelard et al. (2010) reported that the frequency of alleles at the *bg112* locus differed among SSLs. We investigated whether differences in allele frequencies still existed after several generations of in vitro sub-culturing of these SSLs, but using a more conservative approach than the previous study.

Using PCR followed by capillary electrophoresis to distinguish alleles based on their lengths, we found very little variation among four technical replicates of each SSL (Supplementary Table S2), showing that the method is reproducible. Similarly to amplicon sequencing, we found two predominant alleles in C3, which differed by a length of six nucleotides. SSLs shared the same two dominant alleles as C3 (Supplementary Table S2). The ratio of the two alleles in C3 followed a 50:50 ratio (Fig. 3), indicating that this isolate harbours two different nucleotypes in equal frequencies, similarly to its clone A4 (Ropars et al. 2016). Among SSLs, the frequency of allele 1 ranged from 41 to 66% (Fig. 3,

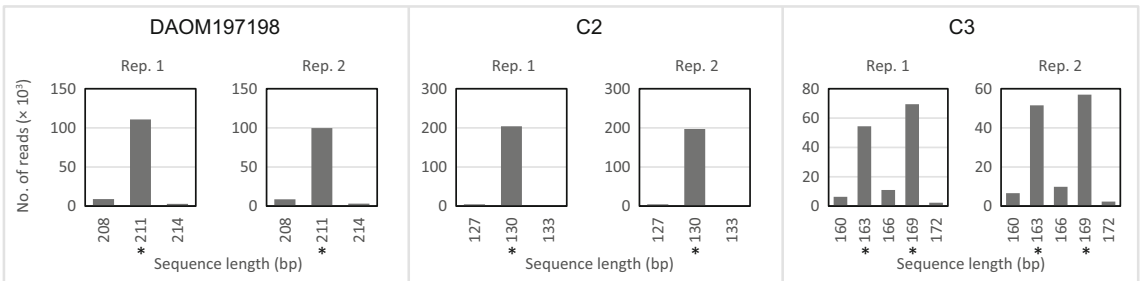

Fig. 2 Number of alleles determined by amplicon sequencing in three *R. irregularis* isolates. *bg112* sequences were amplified using primers Bg112-up and Bg112-d and full amplicons were sequenced with Illumina MiSeq. The number of reads for each sequence was counted and the sequences were sorted by occurrence. Only the predominant sequences named with their respective size are reported on this graph. The number of sequencing reads (occurrence) is reported for each sequence length. Two independent replicates ("Rep. 1" and "Rep. 2") were analysed for the isolates DAOM197198, C2 and C3. The sequences labelled with * are considered true alleles while the other sequences are considered artefacts generated during PCR (stutter effect)

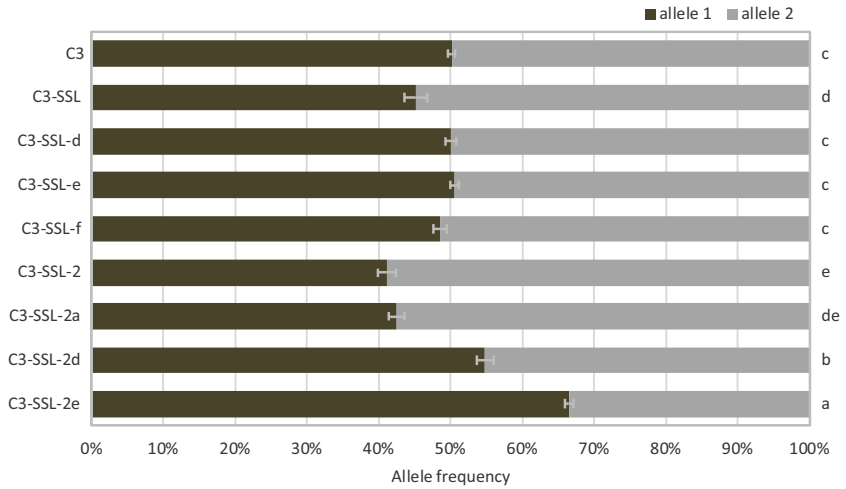

Fig. 3 Relative proportions of two *bg112* alleles in C3 and single-spore lines (SSLs). The allele proportions were measured by PCR followed by capillary electrophoresis. This method measures fluorescence peak height for each sequence length. Only the two predominant peaks corresponding to the identified alleles were retained to calculate the allele proportions. Error bars represent standard deviations ($n = 4$). Bars labelled with the same letter do not differ significantly at $P \leq 0.001$ (general linear model ANOVA, followed by post hoc pair-wise comparisons (Tukey's HSD test)

Supplementary Table S2). Significant differences in allele frequencies were found among the SSLs and the parental isolate C3 (Fig. 3; supplementary results).

Discussion

Single-nucleus sequencing suggested that the *bg112* locus was present in more than one copy in the genome of the *Rhizophagus irregularis* isolate DAOM197198 (Lin et al. 2014), but this finding was based on BLAST searches that identified sequences similar to those sequenced by Angelard et al. (2010). We showed that the *bg112* marker system, defined by Angelard et al. (2010) as being amplified by a precise pair of primers, is indeed a single-copy locus in several isolates of *R. irregularis*. Therefore, this marker can be useful to determine whether or not differences in allele frequencies occur among AMF isolates generated by single-spore propagation and whether or not there are mechanisms maintaining their relative proportions. We propose that in order to better understand the mechanisms of inheritance of nuclei in spores of dikaryotic *R. irregularis* isolates, the approach of investigating allele frequencies of the *bg112* marker and other bi-allelic markers should be used on a large number of SSLs. Analysing the distribution of the allele frequencies in a large number of SSLs will show whether there are limits to unequal inheritance of the two nucleus genotypes and whether most isolates follows a 50:50 proportion or not. Such a finding would be evidence that mechanisms exist ensuring that nucleus proportions do not become heavily skewed during spore formation.

The number of alleles we found in isolates DAOM197198, C2 and C3 reflects their genetic organisation. Indeed, DAOM197198 and C2 each harboured one allele, while C3 harboured two different alleles. C3 is a genetic clone of the dikaryotic isolate A4 (Wyss et al. 2016; Ropars et al. 2016). No allele was shared by all three isolates, reflecting that they belong to different genetic groups (Savary et al. 2018). We detected the two alleles originating from C3 in C3-SSL, C3-SSL-2 and all the derived lines. The allele frequency ratio in C3-SSL and C3-SSL-2, as well as in three SSLs derived from C3-SSL-2, however, deviated from the ratio of C3. Based on *bg112* alleles, our study indicates that the two nucleotypes show unequal frequencies among samples of a clonal dikaryotic AMF.

Coenocytic hyphae, also known as non-septate hyphae, give rise to large multi-nucleated spores through asexual reproduction. Each spore inherits a variable number of nuclei, which is related partly to the spore size (Marleau et al. 2011). Some nuclei migrate directly from the mycelium into the spore while some nuclei likely result from mitosis within the spore (Marleau et al. 2011). In a dikaryotic isolate, the former is likely to result in the inheritance of two different nuclei in unequal proportions due to stochastic processes. If multiple spores are used to initiate a new culture, then variation among individual spores could be masked if the average proportion of nuclei in all the spores is 50:50. It has been observed that in syncytia or other multi-nucleated cells in other organisms, the nuclei divide synchronously, but this is unknown in AMF. In *R. irregularis*, it is unclear whether proportions deviating from 50:50 are stable in SSLs over multiple generations. Under standard conditions in the laboratory, however, we have been able to maintain replicate cultures of SSLs without alteration of the proportions among the replicates.

By introducing selected AMF isolates into the soil, mycorrhizal fungi have an important potential application in agriculture (Ceballos et al. 2013). One possibility is to produce and select AMF varieties that strongly promote plant growth. Classical breeding is not easy with AMF, however, because the process of sexual reproduction in these fungi is not well understood and still debated. The generation of differences in the proportions of nuclei via single-spore

propagation is potentially a promising way to generate cultures with different benefits for improved crop performance. The approach described here would be a way of genetically characterising SSLs. To use newly generated isolates in agriculture, however, it will be necessary to determine whether or not allele frequencies of dikaryotic SSLs are stable and how their effect on plants can change as a result of variation in nucleus frequencies. Nevertheless, the results presented here are promising because significant inequality in allele frequency was observed among SSL samples which differed from those of the parent, several years after they initially were established.

Conclusions

We showed that asexually produced siblings of *R. irregularis* were quantitatively different from a parental isolate in terms of the relative frequency of alleles that are likely located in two different nucleotypes. Even if the level of polymorphism found in *R. irregularis* isolates has recently been revised to lower values than previously thought (Wyss et al. 2016; Ropars et al. 2016), we showed here that single spores inherit different proportions of two nucleus genotypes compared to the parental isolate. Our study reports interesting variation that is generated during vegetative propagation that could account for the variation in fungal traits seen among clonally produced siblings. We stress, however, that the generation of quantitatively detectable genetic variation in *R. irregularis* does not exclude the possibility that these fungi also generate much greater genetic variation through sexual reproduction, even though this has not yet been directly observed.

Acknowledgments We are grateful to Jeremy Bonvin and Pawel Rosikiewicz for their help in culturing the biological material. All bioinformatics computations were performed at the Vital-IT (www.vital-it.ch) Center for High Performance Computing of the Swiss Institute of Bioinformatics and all Illumina MiSeq® sequencing was performed at the Genomic Technologies Facility (GTF) of the University of Lausanne. This work was supported by a grant from the Swiss National Science Foundation (number 31003A_162549).

Financial disclosure This research was funded by a Swiss National Science Foundation grant to IRS (3100 3A_162549). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author's contributions Conceived and designed the experiments: FGM, TW, IRS. Produced the biological material: TW, IDMG. Carried out PCR, sequencing libraries and microsatellite analyses: TW, CA. Bioinformatics analyses: FGM. Data analysis: FGM, TW. Wrote the manuscript: FGM, TW, IRS.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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. Curr Biol 20:1216–1221. https://doi.org/10.1016/j.cub.2010.05.031

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477

Bécard G, Fortin JA (1988) Early events of vesicular-arbuscular mycorrhiza formation on Ri T-DNA transformed roots. New Phytol 108: 211–218. https://doi.org/10.1111/j.1469-8137.1988.tb03698.x

Ceballos I, Ruiz M, Fernández C, Peña R, Rodríguez A, Sanders IR (2013) The *in vitro* mass-produced model mycorrhizal fungus, *Rhizophagus irregularis*, significantly increases yields of the globally important food security crop cassava. PLoS One 8:e70633. https://doi.org/10.1371/journal.pone.0070633

Corradi N, Croll D, Colard A, Kuhn G, Ehinger M, Sanders IR (2007) Gene copy number polymorphisms in an arbuscular mycorrhizal fungal population. Appl Environ Microbiol 73:366–369. https://doi.org/10.1128/AEM.01574-06

Croll D, Sanders IR (2009) Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. BMC Evol Biol 9:13. https://doi.org/10.1186/1471-2148-9-13

Croll D, Wille L, Gamper HA, Mathimaran N, Lammers PJ, Corradi N, Sanders IR (2008) Genetic diversity and host plant preferences revealed by simple sequence repeat and mitochondrial markers in a population of the arbuscular mycorrhizal fungus *Glomus intraradices*. New Phytol 178:672–687. https://doi.org/10.1111/j.1469-8137.2008.02381.x

Ehinger MO, Croll D, Koch AM, Sanders IR (2012) Significant genetic and phenotypic changes arising from clonal growth of a single spore of an arbuscular mycorrhizal fungus over multiple generations. New Phytol 196:853–861. https://doi.org/10.1111/j.1469-8137.2012.04278.x

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hothorn T, Bretz F, Westfall P, Heiberger RM, Schuetzenmeister A, Scheibe S, Hothorn MT 2016. Package 'multcomp'. http://multcomp.r-forge.r-project.org/

Koch AM, Kuhn G, Fontanillas P, Fumagalli L, Goudet J, Sanders IR (2004) High genetic variability and low local diversity in a population of arbuscular mycorrhizal fungi. Proc Natl Acad Sci U S A 101: 2369–2374. https://doi.org/10.1073/pnas.0306441101

Kuhn RM, Haussler D, Kent WJ (2013) The UCSC genome browser and associated tools. Brief Bioinform 14:144–161. https://doi.org/10.1093/bib/bbs038

Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DGO, Mu D, Pang E, Cao H, Cha H, Lin T, Zhou Q, Shang Y, Li Y, Sharma T, van Velzen R, de Ruijter N, Aanen DK, Win J, Kamoun S, Bisseling T, Geurts R, Huang S (2014) Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. PLoS Genet 10:e1004078. https://doi.org/10.1371/journal.pgen.1004078

Marleau J, Dalpé Y, St-Arnaud M, Hijri M (2011) Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. BMC Evol Biol 11:51. https://doi.org/10.1186/1471-2148-11-51

Riley R, Corradi N (2013) Searching for clues of sexual reproduction in the genomes of arbuscular mycorrhizal fungi. Fungal Ecol 6:44–49. https://doi.org/10.1016/j.funeco.2012.01.010

Ropars J, Toro KS, Noel J, Pelin A, Charron P, Farinelli L, Marton T, Krüger M, Fuchs J, Brachmann A, Corradi N (2016) Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. Nat Microbiol 1:16033. https://doi.org/10.1038/nmicrobiol.2016.33

Rosikiewicz P, Bonvin J, Sanders IR (2017) Cost-efficient production of in vitro *Rhizophagus irregularis*. Mycorrhiza 27:477–486. https://doi.org/10.1007/s00572-017-0763-2

Sanders IR, Croll D (2010) Arbuscular mycorrhiza: the challenge to understand the genetics of the fungal partner. Annu Rev Genet 44:271–292. https://doi.org/10.1146/annurev-genet-102108-134239

Savary R, Masclaux FG, Wyss T, Droh G, Cruz Corella J, Machado AP, Morton JB, Sanders IR (2018) A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus *Rhizophagus irregularis*. ISME J 12:17–30. https://doi.org/10.1038/ismej.2017.153

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026

Schmieder R, Lim YW, Rohwer F, Edwards R (2010) TagCleaner: identification and removal of tag sequences from genomic and metagenomic datasets. BMC Bioinformatics 11:341. https://doi.org/10.1186/1471-2105-11-341

Smith SE, Read DJ (2008) Mycorrhizal Symbiosis, 3rd edn. Elsevier Ltd., London

Tisserant E, Kohler A, Dozolme-Seddas P et al (2012) The transcriptome of the arbuscular mycorrhizal fungus Glomus intraradices (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. New Phytol 193:755–769. https://doi.org/10.1111/j.1469-8137.2011.03948.x

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, Frei dit Frey N, Gianinazzi-Pearson V, Gilbert LB, Handa Y, Herr JR, Hijri M, Koul R, Kawaguchi M, Krajinski F, Lammers PJ, Masclaux FG, Murat C, Morin E, Ndikumana S, Pagni M, Petitpierre D, Requena N, Rosikiewicz P, Riley R, Saito K, San Clemente H, Shapiro H, van Tuinen D, Becard G, Bonfante P, Paszkowski U, Shachar-Hill YY, Tuskan GA, Young JPW, Sanders IR, Henrissat B, Rensing SA, Grigoriev IV, Corradi N, Roux C, Martin F (2013) Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proc Natl Acad Sci U S A 110:20117–20122. https://doi.org/10.1073/pnas.1313452110

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

Wyss T, Masclaux FG, Rosikiewicz P et al (2016) Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus *Rhizophagus irregularis*. ISME J:1–13. https://doi.org/10.1038/ismej.2016.29