JOURNAL OF Evolutionary Biology

doi:10.1111/j.1420-9101.2010.02019.x

Intra-isolate genome variation in arbuscular mycorrhizal fungi persists in the transcriptome

E. BOON*, E. ZIMMERMAN*, B. F. LANG† & M. HIJRI*

*Département de Sciences Biologiques, Institut de Recherche en Biologie Végétale, Université de Montréal, Montréal, QC, Canada
†Département de Biochimie, Robert Cedergren Centre for Bioinformatics and Genomics, Université de Montréal, Montréal, QC, Canada

Keywords:

arbuscular mycorrhizal fungi;
Glomus etunicatum;
Glomus intraradices;
intra-organismal genetic heterogeneity;
nuclear variation;
ribosomal variation;
transcriptome variation.

Abstract

Arbuscular mycorrhizal fungi (AMF) are heterokaryotes with an unusual genetic makeup. Substantial genetic variation occurs among nuclei within a single mycelium or isolate. AMF reproduce through spores that contain varying fractions of this heterogeneous population of nuclei. It is not clear whether this genetic variation on the genome level actually contributes to the AMF phenotype. To investigate the extent to which polymorphisms in nuclear genes are transcribed, we analysed the intra-isolate genomic and cDNA sequence variation of two genes, the large subunit ribosomal RNA (LSU rDNA) of *Glomus* sp. DAOM-197198 (previously known as *G. intraradices*) and the *POL1*-like sequence (PLS) of *Glomus etunicatum*. For both genes, we find high sequence variation at the genome and transcriptome level. Reconstruction of LSU rDNA secondary structure shows that all variants are functional. Patterns of PLS sequence polymorphism indicate that there is one functional gene copy, PLS2, which is preferentially transcribed, and one gene copy, PLS1, which is a pseudogene. This is the first study that investigates AMF intra-isolate variation at the transcriptome level. In conclusion, it is possible that, in AMF, multiple nuclear genomes contribute to a single phenotype.

Introduction

Arbuscular mycorrhizal fungi (AMF) are obligate symbionts of plant roots that improve host nutrient uptake and pathogen resistance (Smith & Read, 2008). Plant growth response and productivity vary significantly with the composition of AMF communities (van der Heijden *et al.*, 1998), and the effects of AMF on plant performance are studied intensively. It is estimated that 60–80% of all land plants associate with AMF (Wang & Qiu, 2006; Smith & Read, 2008). The earliest fossils that resemble AMF spores are ~460 million years old, and various authors suggest that AMF may have played an essential role in the establishment of early land plants (Simon *et al.*, 1993; Remy *et al.*, 1994; Redecker *et al.*, 2000).

Despite the obvious importance of AMF to plant functioning and ongoing efforts to understand this symbiosis, the magnitude, organization and expression of genetic variation in AMF are poorly understood. The study of genetic variation in these organisms is unusually complex owing to their unique lifestyle and genome structure. First, from a technical perspective, AMF are difficult to handle because they are obligate biotrophs that are only cultivable in the presence of a host plant. It is therefore challenging to obtain sufficient amounts of contaminant-free nucleic acids for genetic analysis (Hijri *et al.*, 2002). An *in vitro* culture system using transformed carrot roots has been developed to alleviate these problems (St-Arnaud *et al.*, 1996). Using this system, tissue cultures (or 'isolates') are started from single spores and can be maintained under sterile conditions. Second, the study of AMF genetics is conceptually challenging because of the large amount of genetic variation occurring within a single mycelium. AMF are coenocytic; that is, there are no septa between cells, and organelles freely

Correspondence: Eva Boon, Département de Sciences Biologiques, Institut de Recherche en Biologie Végétale, Université de Montréal, 4101 rue Sherbrooke Est, Montréal, QC H1X 2B2, Canada.
Tel.: +1 514 868 5136; fax: +1 514 872 9406;
e-mail: eva.boon@umontreal.ca
Data deposition: Nucleotide sequences were deposited in GenBank under accession numbers FJ235536–FJ235623, GU992000–GU992199 and FJ743698–FJ743715.

move through the hyphal network. The partition of this genetic variation between nuclei has been a subject of controversy, featuring two competing hypotheses.

The first hypothesis states that genetic variation is contained within genetically identical polyploid nuclei (Pawlowska & Taylor, 2004, 2005; Pawlowska, 2005). The second hypothesis describes nuclei as being haploid and heterokaryotic (Kuhn *et al.*, 2001; Hijri & Sanders, 2005). Here, heterokaryosis refers to the co-existence of many genetically differentiated nuclei within the same cytoplasm. The evidence in favour of the first hypothesis is contestable (Bever & Wang, 2005), and recent studies favour the latter hypothesis (Corradi *et al.*, 2007; Angelar *et al.*, 2009; Cárdenas-Flores *et al.*, 2010). In this study, we assume the hypothesis of heterokaryosis.

In multicellular organisms, it is generally assumed that all cells share the same (nuclear) genome. Typically, this results from the transmission of a single progenitor genome, in the form of a spore or gamete, from parent to offspring. However, in organisms that do not pass through a single-genome stage, intra-organismal variation is inevitable and may even play an important role in clonal propagation (see Pineda-Krch & Lehtila, 2004; for a review). For example, heteroplasmy has been reported in both mitochondria (Rand, 2001) and chloroplasts (Petit & Vendramin, 2007), because cytoplasmic genomes do not necessarily pass through a single-copy stage in their life cycle. The potential for intra-organismal variation is closely linked to the magnitude of the genetic bottleneck an organism experiences during its life cycle. To our knowledge, AMF are the only organisms that are multi-nucleate at all life stages, never passing through a uninuclear phase (J. Marleau, Y. Dalpé, M. St-Arnaud & M. Hijri, unpublished).

Intra-isolate variation in AMF has been reported for ribosomal RNA-encoding DNA (rDNA) (Sanders *et al.*, 1995; Clapp *et al.*, 1999, 2001; Pringle *et al.*, 2000; Kuhn *et al.*, 2001; Rodriguez *et al.*, 2001) and for protein-coding genes (Kuhn *et al.*, 2001; Helgason *et al.*, 2003; Corradi *et al.*, 2004, 2007; Corradi & Sanders, 2006). Even copy number polymorphism between isolates has been observed (Corradi *et al.*, 2007). Although both ribosomal and protein-coding genes of AMF are known to possess significant amounts of intra-isolate variation, it is not known whether this intra-isolate variation persists at the transcriptional level. This is an important step towards understanding how a heterokaryotic population of nuclei gives rise to the AMF phenotype.

Why is it possible that the genetic variation we observe among nuclei is not expressed? First, the clonal mode of AMF reproduction could mean that recombination in AMF is a rare event (but see Gandolfi *et al.*, 2003; Croll *et al.*, 2009, Croll & Sanders, 2009). If recombination is rare, the failure to purge deleterious mutations could lead to the presence of pseudogenes. This process is known as Muller's ratchet (Muller, 1964). Second, if multiple gene copies within the AMF mycelium (partitioned either within or between nuclei) behave as genes after a duplication event, they can lose functionality over time (Zhang, 2003). For these two reasons, it is possible that the internuclear or intergenomic variation in AMF is largely comprised of pseudogenes, which are not transcribed and do not contribute to the phenotype.

The extent and organization of genetic variation between nuclei within AMF isolates has thus far been studied using two distinct approaches. First, the presence of different alleles, and variation in copy number between nuclei of a single isolate, has been visualized using FISH techniques for ribosomal (ITS, Kuhn *et al.*, 2001) and protein-coding loci (BiP, Kuhn, 2003). Second, by comparing copy number and allele number, it is possible to infer whether variation is partitioned between or within genomes of a given AMF isolate. The protein-coding gene *POL1*-like sequence (PLS) was studied using this latter approach (Pawlowska & Taylor, 2004; Hijri & Sanders, 2005). Two loci that vary between nuclei within the same isolate have been studied in detail: large subunit ribosomal DNA (LSU rDNA) and PLS. In this study, we compare the genetic diversity of these latter loci at the genome level with the genetic diversity at the transcriptome level. More specifically, we compare the degree of within-isolate nucleotide polymorphism in genomic versus complementary DNA (gDNA vs. cDNA), for *Glomus* sp. (DAOM-197198) LSU rDNA and *Glomus etunicatum* PLS genes. If we find less variation on the transcriptome level, it would support the hypothesis that a portion of AMF genetic diversity is not expressed and raise the possibility that much of AMF intra-isolate diversity consists of pseudogenes. We address two specific questions: (i) are gene variants transcribed? and (ii) is there any evidence that alleles are under relaxed selective pressure?

Materials and methods

DNA extraction and cDNA library construction

DNA and RNA were extracted with the Qiagen Plant DNA extraction kit (Qiagen, Toronto, ON, Canada) and RNAqueous®-Micro (Ambion, Streetsville, ON, Canada), from freshly harvested spores and hyphae of *Glomus* sp. isolate DAOM-197198, previously known as *G. intraradices* (Stockinger *et al.*, 2009) and *Glomus etunicatum* isolate Native Plants Incorporated (NPI). We did not use the same species for both markers, because thus far, PLS alleles have never been successfully amplified in *Glomus* sp. isolate DAOM-197198. cDNA synthesis was performed using RevertAid™ First Strand cDNA synthesis kits (Fermentas, Burlington, ON, Canada) and tested for gDNA contamination by comparing PCR fragment sizes of the rDNA intergenic spacer (IGS) and of (intron-containing) P-Type IID ATPases (Corradi & Sanders, 2006). No gDNA contamination was detected in the cDNA libraries.

DNA amplification

The entire LSU rDNA was amplified from gDNA with the forward primer F1.LSU and the reverse primer R1.LSU (see Table 1 for primer sequences), yielding fragments of 3056–3091 bp. An approximately 200-bp region was selected for further investigation and amplified from gDNA and cDNA using the same forward primer F1.LSU and the reverse primer R1.LSU.cDNA (see Table 1). This section covers a region that contains both highly variable and conserved sections and that corresponds to structure **a** in Figure 1(a) in Schnare *et al.* (1996). *Glomus etunicatum* cDNA was amplified by PCR using the primers Pol4 and Pol7 (Pawlowska & Taylor, 2004). High fidelity *Pfu* DNA polymerase (Fermentas) was used for all PCR, under standard conditions, with 35 cycles. Cloning was performed with the TOPO® TA Cloning Kit (Invitrogen, Carlsbad, CA, USA), and sequencing with universal primers was carried out at the Genome Quebec Innovation Centre sequencing platform (Montreal, Quebec).

Testing for *in vitro* recombination and polymerase errors

If PCR is performed on a mixed template, *in vitro* recombination can bias polymorphism estimates. This is especially relevant when using *Pfu* polymerases and a higher number of PCR cycles (between 20 and 40) (Zylstra *et al.*, 1998). The heterokaryotic state of the AMF under study would provide a mixed template for the PCR, and we thus evaluated the role of *in vitro* recombination on polymorphism in our dataset. We performed a PCR on a mixed template of equivalent amounts of two plasmids containing highly divergent PLS sequences (belonging to the groups PLS1 and PLS2), which served as our reference sequences, under the same conditions as those used for our main PLS dataset (gDNA and cDNA). We subsequently cloned and sequenced the PCR products and compared the results with the initial two reference sequences. We recovered six PLS1 sequences and eight PLS2 sequences. No recombination events and no polymerase errors were observed.

Table 1 Primers used for LSU rDNA and PLS amplification.

Primer name	Sense	5′–3′ sequence
Pol4-A Titanium adaptor sequence	Forward	GAATCCTTCCCAAATTGATCAGAATACTTGTT CCATCTCATCCCTGCGTGTCTCCGACTCAG
Pol7-B Titanium adaptor sequence	Reverse	TAATAATAAAAGCCTTTCAAAAAATCCATCAATA CCTATCCCCTGTGTGCCTTGGCAGTCTCAG
F1.LSU	Forward	GCATATCAATAAGCGGAGGA
R1.LSU	Reverse	CGGTCTAAACCCAGCTCACG
R1.LSU. cDNA	Reverse	CTAATAGGGAACGTGAGCT

PLS pyrosequencing

In an attempt to exhaustively sample PLS allelic diversity, we sequenced a single *G. etunicatum* spore using 454 technology. A single spore was picked, placed in 1 μL sterile water and crushed on the bottom of 0.2-mL tubes using a Pasteur pipette on which the tip had been melted into a ball. DNA of this spore was amplified using the GenomiPhi Whole-Genome Amplification kit (GE Healthcare, Amersham, UK) according to the manufacturer's instructions. PLS variants were amplified using Dream*Taq* DNA polymerase (Fermentas) using primers Pol4 and Pol7, with added Titanium adaptor sequences for pyrosequencing. The reaction was performed in 50 μL volumes containing 0.2 mM dNTPs, 0.5 μmol of each primer and the PCR buffer. PCR was carried out for 35 cycles (95 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min; preceded by an initial 2-min denaturation at 90 °C and followed by a 10-min hold at 72 °C) on a Mastercycler epgradient S (Eppendorf, Mississauga, ON, Canada). The PCR product was loaded on an electrophoresis gel to ensure successful amplification of the gene, and the bands were cut from the gel and purified by 24-h incubation in 50 μL milliQ water. Purified samples were sent to the Genome Quebec Innovation Centre in Montreal for pyrosequencing using the GS FLX Titanium emPCR kit (Lib-A), with unidirectional reads (1/8 run per sample).

Sequence alignment and data analyses

Sequence alignments for both the LSU rDNA and PLS loci were performed with MAFFT (Katoh *et al.*, 2005) in the Jalview interface (Clamp *et al.*, 2004) and then verified and refined by eye using Bioedit v7.0.5 (Hall, 1999). For each sequence, indel composition was registered before excluding alignment gaps from further statistical analysis.

We used DNAsp v4.50.3 (Rozas *et al.*, 2003) for polymorphism and genetic differentiation analyses and calculated a Principal Component Analysis matrix in Jalview. Neighbour-joining trees were inferred using MEGA v.4.0 (Tamura *et al.*, 2007). A Neighbour-Net network (uncorrected *P*-distance) was constructed for the PLS locus using SplitsTree4 (Huson & Bryant, 2006), testing statistical significance with bootstrap analyses (1000 replicates).

rDNA secondary structure was inferred by aligning LSU sequences of *Glomus* sp. (isolate DAOM-197198), with the secondary structures of *Tricholoma matsutake* (Hwang & Kim, 2000) and *Saccharomyces cerevisiae* (U53879) from the Comparative RNA Web website (http://www.rna.ccbb.utexas.edu) as references. Pairings

were identified both by sequence comparison at the primary sequence level and by searching for compensatory base pair mutations with CBCanalyzer (Wolf *et al.*, 2005). Compensatory base pair mutations preserve nucleotide interactions and thus the secondary structure of the ribosome that is formed by these nucleotide interactions. For example, if an 'A' mutates into a 'G', then the 'U' that originally paired with the 'A' will change into a 'C' to preserve base pairing and thus secondary structure.

We investigated substitution patterns in the context of the stem-loop structure of both the complete 2053-bp LSU rDNA alignment and a 220-bp subset. We divided the LSU rRNA gene of *Glomus* sp. isolate DAOM 197198 into functional regions, following the description of the eukaryote LSU rRNA gene in Schnare *et al.* (1996) (see also Table S2). To further examine the effect of nucleotide substitutions on the conservation of secondary structure, we calculated the amount of compensatory base changes (CBC) in these functional regions, using additional sequences from other isolates (additional sequences listed in Table S3).

Results

Sequences obtained

For the LSU rRNA gene, we obtained 14 sequences from gDNA covering most of the gene's 5′ region (2053 bp, accession numbers FJ235561–FJ235574). For the 220-bp subset of this region, we recovered 33 sequences from cDNA (accession numbers FJ235536–FJ235560 and FJ743698–FJ743705) and 24 from gDNA (FJ235561–FJ235574 and FJ743706–FJ743715). Additional sequences were downloaded from GenBank; 18 from gDNA and 22 ESTs (see Supporting Information for accession numbers). One contig (gDNA) was obtained from the *Glomus* Genome Consortium database (contig 95064 of the October 16, 2008 assembly) (Martin *et al.*, 2008). All sequences originated from *Glomus* sp. isolate DAOM 197198 and, once aligned, resulted in two alignments; a 2053-bp alignment with 14 sequences and a subset covering 220 bp with 112 sequences. A secondary structure was estimated covering the entire 5′ end of the large subunit (available upon request).

For the PLS gene, we obtained 30 gDNA sequences by cloning (accession numbers FJ235575–FJ235604) and 19 cDNA sequences (accession numbers FJ235605–FJ235623). Pyrosequencing yielded 1250 sequences, of which the minimum length was 39 bp and the maximum length was 454 bp, with an average length of 130 bp. Of these, 200 sequences longer than 200 bp were retained (NCBI accession numbers GU992000–GU992199). PLS sequences were pooled with 15 previously published sequences (accession numbers AY330581, AY394011–AY394024. All sequences came from *G. etunicatum* isolate NPI (Plantworks, Kent, UK).

LSU rDNA polymorphism and genetic differentiation

Haplotype diversity H_d and pairwise nucleotide diversity π (Nei, 1987) for the LSU rDNA are summarized in Table 2. Both estimators are measures of sequence polymorphism. Pairwise nucleotide diversity π was also calculated along a sliding window for the 2053-bp alignment (Fig. 1).

Genetic differentiation was estimated using several statistical approaches. We used the nearest-neighbour statistic (S_{nn} statistic, Hudson, 2000), which is a sequence-based estimator of particular use for datasets with high nucleotide diversity. The choice of a nonparametric measure of population differentiation such as S_{nn} is justified, because, in the case of AMF, we have very little information about the way genetic variation is generated and maintained. To obtain a second independent estimate, genetic differentiation was also estimated using a principal component analysis (results not shown). For either method, no genetic differentiation was detected within or between the 2053-bp and 220-bp alignments.

LSU rDNA secondary structure

Ribosomal regions A, B, D and H, which are known to be variable and/or dispensable in other species (details listed in Table S2), also show high genetic variation in our alignment (Fig. 1). In general, genetic diversity was higher in regions that formed loops and lower in stem regions. For regions C and F, no CBCs were calculated, because region C interacts with the small subunit of the ribosomal DNA for which pairing sequences were unavailable and region F appears to be a Glomeromycota-specific insertion. No CBC changes were observed for region E.

PLS nucleotide polymorphism and population differentiation

By comparing cDNA and genomic sequences from *G. etunicatum*, we found that the PLS gene contains two introns. Our data were complemented with additional

Table 2 LSU rDNA polymorphism estimates.

Alignment length	# Sequences	# Alleles	H_d*	π†
2053 bp				
Genomic	14	12	0.967	0.01872
220 bp				
All	112	72	0.977	0.05362
Genomic	57	37		
cDNA	55	35		

*Haplotype diversity (Nei, 1987).
†Nucleotide diversity; average number of nucleotide differences per site between two sequences (Nei, 1987).

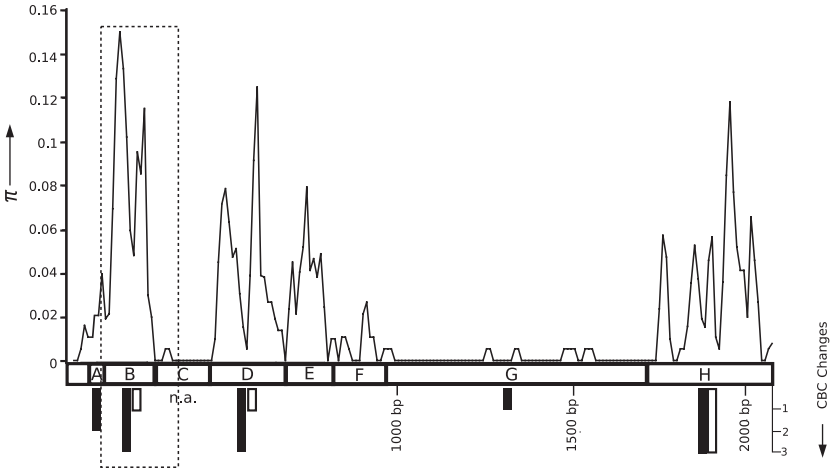

Fig. 1 Within-isolate nucleotide diversity (π) (Nei, 1987) along the 5′ LSU rDNA of *Glomus intraradices* for the entire 2053-bp LSU rDNA, with a comparison of compensatory base pair changes within *Glomus* sp. isolate DAOM-197198 (white bars) and between isolate DAOM-197198 and other *G. intraradices* isolates (black bars). The dotted rectangle indicates the relative location of the 220-bp subset alignment.

sequences from a previously published dataset (Pawlowska & Taylor, 2004), and we observed no genetic differentiation between the two datasets.

At the PLS locus, 65 alleles were retrieved from 263 sequences. These alleles fall into two genetically distinct groups (S_{nn} = 1.000, *P*-value 0.000) that correspond to two previously described gene variants, PLS1 and PLS2 (Pawlowska & Taylor, 2004; Hijri & Sanders, 2005). Patterns of sequence variation and pairwise nucleotide diversity (π) differ markedly between PLS1 and PLS2 sequences. For PLS2, areas of high genetic diversity are limited primarily to introns, whereas in the PLS1 group, exons show extensive polymorphism as well (Table 3). The PLS1 group is more genetically diverse than the PLS2 group (π is 0.01987 and 0.01101, respectively). PLS1 and PLS2 allele frequencies change drastically between sequences that originate from gDNA and cDNA. For PLS1, we found 60 alleles in the gDNA and two in the cDNA, and for PLS2, six and three alleles, respectively (excluding gaps). Finally, 101 of 217 PLS1 gDNA alleles have a deletion in the first exon, representing 65 unique events leading to a frameshift mutation. In two PLS1 alleles, we found a stop codon in the first exon. These observations are summarized in Table 3 and in the Neighbour-Net network of Fig. 2.

Table 3 Polymorphism estimates for PLS (*Glomus etunicatum*).

	n	# Alleles	H_d	π	Frameshift mutations	Stop codon
PLS1						
Genomic						
All	217	60	0.948	0.01987	101†	
Exon				0.01484		2
Intron				0.02738		
cDNA	3	2	0.667	0.02083	3‡	
PLS2						
Genomic						
All	27	6	0.649	0.01101	0	
Exon				0		
Intron				0.01839		
cDNA	16	3	0.242	0.00569	0	

n, number of sequences; H_d, haplotype diversity; pairwise nucleotide diversity (Nei, 1987).
†At least 65 independent mutational events.
‡Three frameshift mutations in same sequence; reading frame not disturbed.

Discussion

High genetic variation persists at the transcript level

The analyses of both LSU rDNA and PLS sequences confirm high intra-isolate genetic polymorphism at the genome level, as shown previously for both the LSU rDNA (Rodriguez *et al.*, 2004) and PLS (Pawlowska & Taylor, 2004). This study is the first to show that genetic variation persists at the transcript level (Tables 2 and 3).

The high levels of intra-isolate sequence polymorphism may raise concerns about the frequency of PCR errors, which could artificially increase sequence variability. However, the reported error rate (2.6×10^{-6} errors per nucleotide per cycle) for *Pfu* polymerase (Lundberg *et al.*, 1991) is much lower than the observed number of nucleotide differences per nucleotide per sequence (π) (Nei, 1987), which was 0.01872 for the ribosomal gene in the *Glomus* sp. isolate DAOM-197198 and 0.01987 and 0.01101 for PLS1 and PLS2 in *G. etunicatum*, respectively. The manufacturer of 454 FLX Titanium technology (Roche) reports an accuracy of 99.5% at 250 bases. Our estimates of nucleotide diversity are quite conservative, considering that the PCR approach with conserved primers we used will retrieve only the most abundant alleles in the isolates under study. Moreover, the sequences retrieved from GenBank were not genetically distinct from our own sequences, suggesting that our dataset is unbiased. Finally, specifically designed tests did not detect *in vitro* recombination or *Pfu* error (see Materials and methods). Considering these arguments, we believe that the genetic diversity we

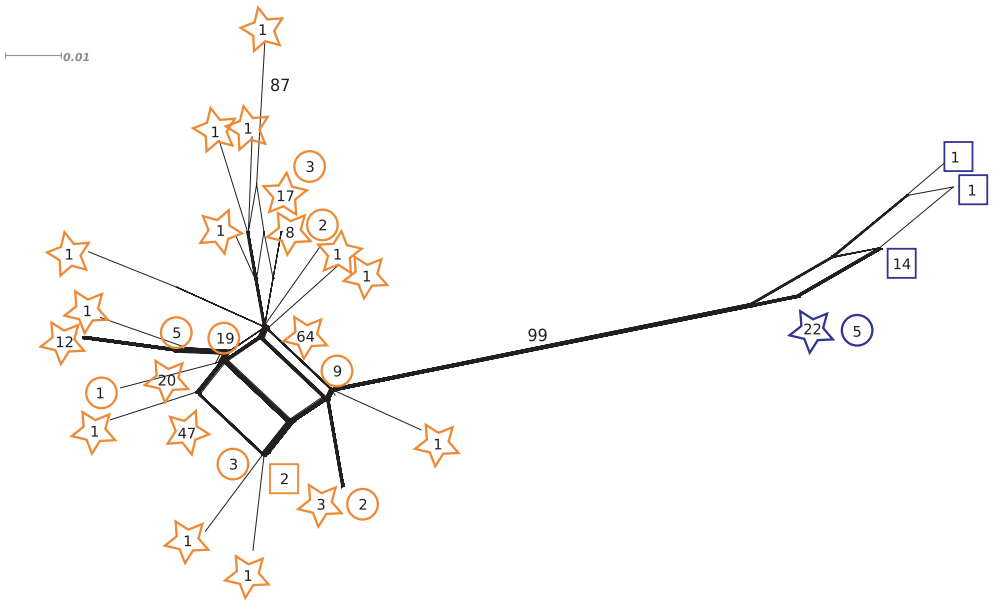

Fig. 2 A Neighbour-Net network for the PLS locus, based on uncorrected *P*-distances. Only exons were taken into account. Reticulation of the branches indicates uncertainties owing to homoplasy (recombination or multiple hits). Sequences that belong to the PLS1 sequence type are represented in orange/light grey (on left), and sequences that belong to the PLS2 sequence type are represented in blue/dark grey (on right). The shape of the symbol corresponds to the mode of sequencing and/or origin of the template; squares are sequences from the cDNA library, circles are gDNA sequences obtained by cloning and stars are gDNA sequences obtained by pyrosequencing. Numbers in the symbols correspond to the frequency of the allele in question. Bootstrap support values >70 are depicted above branches (1000 replicates). Scale bar indicates substitutions per site.

found within AMF isolates is not an artefact. To investigate whether the variant alleles evolve under selection, we explored sequence conservation and patterns of polymorphism for the two loci.

LSU rDNA sequence conservation: a population of sequences under selective constraints

Conserved regions in the *Glomus* sp. LSU rDNA nucleotide sequence correspond well to regions of structural conservation known for *S. cerevisiae* (Schnare *et al.*, 1996) (Fig. 1). Also, regions B and H, which were experimentally shown to be dispensable in yeast (Musters *et al.*, 1989), show high levels of variation in our *Glomus* sp. isolate DAOM 197198. The number of complementary base pair changes (CBC) follows this pattern, because regions B and H have up to three CBCs, whereas the structurally important region G, which contains the GTP-associated centre of the large subunit, only contains a single CBC (Fig. 1). The same pattern of conserved and variable regions is observed for different isolates (see Table S3 for accession numbers).

We found 72 LSU rDNA alleles in the isolate DAOM-197198 (Table 2). To investigate how this variation is partitioned within the cytoplasm, we estimated copy number per genome using absolute quantitative real-time PCR (see Data S1). LSU rDNA copy number is 31.27 ± 0.25 SE, which means that the high number of alleles for LSU rDNA cannot be explained exclusively by polyploidy, as is also the case for the PLS loci (Hijri & Sanders, 2005). This observation lends further support to the hypothesis that nuclei in AMF contain genetically divergent genomes (Kuhn *et al.*, 2001; Hijri & Sanders, 2005).

We are aware that the aforementioned copy number per genome is merely an average over the entire nuclear population within this isolate. It is theoretically possible that copy number per nucleus varies widely and that most variant alleles are still contained within the same nucleus. However, considering that (i) the standard error for our copy number estimate is very low and (ii) concerted evolution of rDNA genes will tend to homogenize gene copies within a genome (Smith, 1976; Arnheim, 1983), it is still unlikely that polyploidy alone can explain our observations. Still, only visualization techniques such as FISH (Fluorescence *In Situ* Hybridization) can conclusively show how genetic variation is partitioned between nuclei. FISH has indeed been performed on two AMF protein-coding genes, demonstrating intra-isolate genetic variation for these two loci (Kuhn *et al.*, 2001; Kuhn, 2003).

PLS groups show contrasting patterns of diversity

We confirm the presence of two genetically divergent groups of PLS alleles, PLS1 and PLS2, that were reported

previously in *G. etunicatum* (Pawlowska & Taylor, 2004; Hijri & Sanders, 2005). Allele frequencies are markedly different between PLS1 and PLS2, both in the genome and in the transcriptome (Table 3 and Fig. 2). We are confident that this difference in allele frequencies between PLS1 and PLS2 reflects the genetic variation in the genome and is not attributable to sampling bias or technical artefacts, because we used two different approaches to obtain our sequences. These approaches include conventional cloning and subsequent Sanger sequencing, as well as massive parallel pyrosequencing using Titanium technology. Both datasets yielded the same skewed allele frequencies.

We propose that the two PLS alleles arose by gene duplication and that the resulting paralogues underwent different evolutionary trajectories, based on the following evidence. First, Hijri & Sanders (2005) proposed that PLS occurs in two copies, based on RT-PCR experiments. Second, genetic diversity is up to four times higher among PLS1-type alleles than among PLS2-type alleles, indicating relaxed selection pressures for PLS1 (Table 3). Third, all PLS2 alleles found in the gDNA were found back in the cDNA, whereas only 3.3% of PLS1 alleles were found back in the cDNA libraries. Finally, 101 frameshift mutations and two stop codons occurred in the PLS1 gDNA alleles. We propose that PLS2 represents the functional gene copy, whereas PLS1 shows signs of relaxed selection pressures and may in fact be a pseudogene.

Concluding remarks

Genetic variation among AMF nuclei within a mycelium has been demonstrated at a genomic level (Hijri & Sanders, 2005). This study is the first to show that this variation is transcribed. Our work demonstrates the presence of many LSU rDNA alleles within a single AMF isolate, all of which show structural conservation. We present evidence for a duplication of the protein-coding gene PLS, of which one duplicate shows signs of relaxed selective pressures. For PLS, genetic variation at the genome level is not necessarily representative of genetic variation in the transcriptome. Our findings are consistent with previous observations of large amounts of genetic variation for protein-coding genes (Hijri & Sanders, 2005), ITS sequences (Hijri *et al.*, 1999; Kuhn *et al.*, 2001) and noncoding sequences (Hijri *et al.*, 2007) within the genome of *Glomus* sp. isolate DAOM-197198 (Martin *et al.*, 2008). The question remains as to how these unique organisms function with tens to hundreds of divergent alleles within a single mycelium.

Acknowledgments

This work was supported by a grant of the Natural Sciences and Engineering Research Council of Canada (NSERC) to MH and a Vanier Canada Graduate Scholarship to EB. The *G. etunicatum* isolate NPI was kindly provided by PlantWorks Ltd, Kent (UK). We thank the Joint Genome Institute and the Glomus Genome Consortium for access to the Glomus genome sequence before publication. Taïca Replansky and Eric Bapteste provided insightful comments on this manuscript.

References

Angelar, C., Croll, D. & Sanders, I.R. (2009) Alteration of AMF genotypes and phenotypes caused by segregation and by change in host species. In: *6th International Conference on Mycorrhyza*. pp. 29. Belo Horizonte, Brazil.

Arnheim, N. (1983) Concerted evolution of multigene types. In: *Evolution of Genes and Proteins* (Sinauer, ed.). pp. 38–61. Sinauer, Sanderland.

Bever, J.D. & Wang, M. 2005. Arbuscular mycorrhizal fungi – hyphal fusion and multigenomic structure. *Nature* **433**: E3–E4.

Cárdenas-Flores, A., Draye, X., Bivort, C., Cranenbrouck, S. & Declerck, S. 2010. Impact of multispores in vitro subcultivation of Glomus sp. MUCL 43194 (DAOM 197198) on vegetative compatibility and genetic diversity detected by AFLP. *Mycorrhiza*, doi: 10.1007/500572-009-0295-5.

Clamp, M., Cuff, J., Searle, S.M. & Barton, G.J. 2004. The Jalview Java alignment editor. *Bioinformatics* **20**: 426–427.

Clapp, J.P., Fitter, A.H. & Young, J.P.W. 1999. Ribosomal small subunit sequence variation within spores of an arbuscular mycorrhizal fungus, Scutellospora sp. *Mol. Ecol.* **8**: 915–921.

Clapp, J.P., Rodriguez, A. & Dodd, J.C. 2001. Inter- and intra-isolate rRNA large subunit variation in Glomus coronatum spores. *New Phytol.* **149**: 539–554.

Corradi, N. & Sanders, I.R. 2006. Evolution of the P-type II ATPase gene family in the fungi and presence of structural genomic changes among isolates of Glomus intraradices. *BMC Evol. Biol.* **6**: 21.

Corradi, N., Kuhn, G. & Sanders, I.R. 2004. Monophyly of beta-tubulin and H+-ATPase gene variants in Glomus intraradices: consequences for molecular evolutionary studies of AM fungal genes. *Fungal Genet. Biol.* **41**: 262–273.

Corradi, N., Croll, D., Colard, A., Kuhn, G., Ehinger, M. & Sanders, I.R. 2007. Gene copy number polymorphisms in an arbuscular mycorrhizal fungal population. *Appl. Environ. Microbiol.* **73**: 366–369.

Croll, D. & Sanders, I. 2009. Recombination in Glomus intraradices, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC Evol. Biol.* **9**: 13.

Croll, D., Giovannetti, M., Koch, A.M., Sbrana, C., Ehinger, M., Lammers, P.J. & Sanders, I.R. 2009. Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytol.* **181**: 924–937.

Gandolfi, A., Sanders, I.R., Rossi, V. & Menozzi, P. 2003. Evidence of recombination in putative ancient asexuals. *Mol. Biol. Evol.* **20**: 754–761.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic Acids Symposium Series*, Vol. 41. pp. 95–98. Oxford University Press, Oxford.

van der Heijden, M.G.A., Klironomos, J.N., Ursic, M., Moutoglis, P., Streitwolf-Engel, R., Boller, T., Wiemken, A. & Sanders, I.R. 1998. Mycorrhizal fungal diversity determines plant

biodiversity, ecosystem variability and productivity. *Nature* **396**: 69–72.

Helgason, T., Watson, I.J. & Young, J.P.W. 2003. Phylogeny of the Glomerales and diversisporales (Fungi : Glomeromycota) from actin and elongation factor 1-alpha sequences. *FEMS Microbiol. Lett.* **229**: 127–132.

Hijri, M. & Sanders, I.R. 2005. Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei. *Nature* **433**: 160–163.

Hijri, M., Hosny, M., van Tuinen, D. & Dulieu, H. 1999. Intraspecific ITS polymorphism in Scutellospora castanea (Glomales, Zygomycota) is structured within multinucleate spores. *Fungal Genet. Biol.* **26**: 141–151.

Hijri, M., Redecker, D., Petetot, J., Voigt, K., Wostemeyer, J. & Sanders, I.R. 2002. Identification and isolation of two ascomycete fungi from spores of the arbuscular mycorrhizal fungus Scutellospora castanea. *Appl. Environ. Microbiol.* **68**: 4567–4573.

Hijri, M., Niculita, H. & Sanders, I.R. 2007. Molecular characterization of chromosome termini of the arbuscular mycorrhizal fungus Glomus intraradices (Glomeromycota). *Fungal Genet. Biol.* **44**: 1380–1386.

Hudson, R.R. 2000. A new statistic for detecting genetic differentiation. *Genetics* **155**: 2011–2014.

Huson, D.H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**: 254–267.

Hwang, S.K. & Kim, J.G. 2000. Secondary structural and phylogenetic implications of nuclear large subunit ribosomal RNA in the ectomycorrhizal fungus Tricholoma matsutake. *Current Microbiology* **40**: 250–256.

Katoh, K., Kuma, K., Toh, H. & Miyata, T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**: 511–518.

Kuhn, G. 2003. *Organisation of Genetic Variation in Multinucleate Arbuscular Mycorrhizal Fungi*. University of Lausanne, Lausanne.

Kuhn, G., Hijri, M. & Sanders, I.R. 2001. Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature* **414**: 745–748.

Lundberg, K.S., Shoemaker, D.D., Adams, M.W.W., Short, J.M., Sorge, J.A. & Mathur, E.J. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* **108**: 1–6.

Martin, F., Gianinazzi-Pearson, V., Hijri, M., Lammers, P., Requena, N., Sanders, I.R., Shachar-Hill, Y., Shapiro, H., Tuskan, G.A. & Young, J.P.W. 2008. The long hard road to a completed Glomus intraradices genome. *New Phytol.* **180**: 747–750.

Muller, H.J. 1964. The relation of recombination to mutational advantage. *Mutat. Res.* **1**: 2–9.

Musters, W., Venema, J., Vanderlinden, G., Vanheerikhuizen, H., Klootwijk, J. & Planta, R.J. 1989. a system for the analysis of yeast ribosomal DNA mutations. *Mol. Cell. Biol.* **9**: 551–559.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Pawlowska, T.E. 2005. Genetic processes in arbuscular mycorrhizal fungi. *FEMS Microbiol. Lett.* **251**: 185–192.

Pawlowska, T.E. & Taylor, J.W. 2004. Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature* **427**: 733–737.

Pawlowska, T.E. & Taylor, J.W. 2005. Arbuscular mycorrhizal fungi – hyphal fusion and multigenomic structure – reply. *Nature* **433**: E4.

Petit, R. & Vendramin, G.G. 2007. Plant phylogeography based on organelle genes: an introduction. In: *Phylogeography of Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity* (S. Weiss & N. Ferrand, eds), pp. 23–97. Springer, Heidelberg.

Pineda-Krch, M. & Lehtila, K. 2004. Costs and benefits of genetic heterogeneity within organisms. *J. Evol. Biol.* **17**: 1167–1177.

Pringle, A., Moncalvo, J.M. & Vilgalys, R. 2000. High levels of variation in ribosomal DNA sequences within and among spores of a natural population of the arbuscular mycorrhizal fungus Acaulospora colossica. *Mycologia* **92**: 259–268.

Rand, D.M. 2001. The units of selection on mitochondrial DNA. *Annu. Rev. Ecol. Syst.* **32**: 415–448.

Redecker, D., Kodner, R. & Graham, L.E. 2000. Glomalean fungi from the Ordovician. *Science* **289**: 1920–1921.

Remy, W., Taylor, T.N., Hass, H. & Kerp, H. 1994. 4-Hundred-million-year-old vesicular-arbuscular mycorrhizae. *Proc. Natl Acad. Sci. USA* **91**: 11841–11843.

Rodriguez, A., Dougall, T., Dodd, J.C. & Clapp, J.P. 2001. The large subunit ribosomal RNA genes of Entrophospora infrequens comprise sequences related to two different glomalean families. *New Phytol.* **152**: 159–167.

Rodriguez, A., Clapp, J.P. & Dodd, J.C. 2004. Ribosomal RNA gene sequence diversity in arbuscular mycorrhizal fungi (Glomeromycota). *J. Ecol.* **92**: 986–989.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X. & Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.

Sanders, I.R., Alt, M., Groppe, K., Boller, T. & Wiemken, A. 1995. Identification of ribosomal DNA polymorphisms among and within Spores of the glomales – application to studies on the genetic diversity of arbuscular mycorrhizal fungal communities. *New Phytol.* **130**: 419–427.

Schnare, M.N., Damberger, S.H., Gray, M.W. & Gutell, R.R. 1996. Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23 S-like) ribosomal RNA. *J. Mol. Biol.* **256**: 701–719.

Simon, L., Bousquet, J., Levesque, R.C. & Lalonde, M. 1993. Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. *Nature* **363**: 67–69.

Smith, G.P. 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* **141**: 528–534.

Smith, S.E. & Read, D.J. 2008. *Mycorrhizal Symbiosis*, 3rd edn. Academic Press, London.

St-Arnaud, M., Hamel, C., Vimard, B., Caron, M. & Fortin, J.A. 1996. Enhanced hyphal growth and spore production of the arbuscular mycorrhizal fungus Glomus intraradices in an in vitro system in the absence of host roots. *Mycol. Res.* **100**: 328–332.

Stockinger, H., Walker, C. & Schüssler, A. 2009. '*Glomus intraradices* DAOM197198', a model fungus in arbuscular mycorrhiza research, is not *Glomus intraradices*. *New Phytol.* **183**: 1176–1187.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**: 1596–1599.

Wang, B. & Qiu, Y.L. 2006. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**: 299–363.

Wolf, M., Friedrich, J., Dandekar, T. & Müller, T. 2005. CBCAnalyzer: inferring phylogenies based on compensatory

base changes in RNA secondary structures. *In Silico Biol.* **5**: 0027.
Zhang, J.Z. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**: 292–298.
Zylstra, P., Rothenfluh, H.S., Weiller, G.F., Blanden, R.V. & Steele, E.J. 1998. PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. *Immunol. Cell Biol.* **76**: 395–405.

Supporting information

Additional Supporting Information may be found in the online version of this article:
Table S1 Additional LSU sequences from Genbank.
Table S2 Description of LSU regions.
Table S3 Additional genomic sequences used for inter-isolate comparison in the LSU rDNA analysis.
Data S1 Absolute quantification of LSU rDNA copy number.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received 2 December 2009; revised 8 April 2010; accepted 12 April 2010