

Letters

Multilocus genotyping of arbuscular mycorrhizal fungi and marker suitability for population genetics

Arbuscular mycorrhizal fungi (AMF) are an ecologically important group of plant symbionts and their species richness has been shown to influence plant diversity and productivity (Van der Heijden *et al.*, 1998). Genetic diversity within AMF species is important as genetically different isolates have been shown to differentially affect plant growth and nutrition (Munkvold *et al.*, 2004; Koch *et al.*, 2006). The study of AMF diversity in ecosystems, particularly identifying which AMF species associate with different host plants, requires reliable identification of different AMF. It has long been recognized that identifying AMF across broad geographical ranges requires molecular tools for fast and reliable genotyping directly from soil material.

Previously, genotyping methods for distinguishing AMF species have mostly been restricted to ribosomal DNA (rDNA) sequences. The advantage of these loci is the potential for cross-species amplification using universal primers, and the relative ease of amplification from different material (e.g. colonized root pieces, single spores, etc.). A large body of studies have identified the species composition of AMF communities in many different ecosystems (Öpik *et al.*, 2006; Rosendahl, 2008). However, studies of genetic variability within AMF species are important for understanding the basic biology, genetics and ecology of AMF fungi, which cannot be addressed at the community level. For example, a hierarchical study of genetic variability from the local scale within populations right up to an inter-continental scale is lacking. Such hierarchically designed studies could lay the foundation that will allow us to answer fundamental questions about the biology of AMF, their genetics, whether they form recombinant populations, the amount of genetic exchange among populations, the importance of drift and selection in AMF species, and the distribution of genetic and functional diversity in AMF over different geographic scales, and allow us to examine the co-evolutionary relationships between AMF genotypes and their host plant genotypes.

For most of these applications ribosomal markers are unsuitable because of a lack of sufficient within-species variability and are potentially problematic because of confounding intra-sporal variability (Sanders *et al.*, 1995) and

copy number polymorphism (Corradi *et al.*, 2007). A population genetics approach to the study of AMF requires multilocus genotyping of nonribosomal loci. Stukenbrock & Rosendahl (2005a,b) first developed and applied this approach by amplifying three different loci in a large set of spores of three *Glomus* species harvested from the field. However, ideally, multilocus genotyping should comprise a much larger number of loci. Two simultaneously published studies (Croll *et al.*, 2008; Mathimaran *et al.*, 2008), describing genetic markers for AMF, should now make this possible. Both studies identified multiple loci that were variable among isolates of a commonly studied AMF, *Glomus intraradices*. Length differences among the alleles were used to identify genetic differences. Part, but not all, of the variation was found in repeat regions, and both studies referred to the markers as either microsatellites or simple sequence repeat markers. The simultaneous publication of the two studies might lead to some confusion for researchers who may now want to use these markers. Here, our aim is to clarify how many new and different loci have actually been identified and which loci are likely to be suitable for population genetics studies, to highlight potential problems with the genotyping techniques used, and to discuss future approaches to their use in AMF population biology.

The study by Mathimaran *et al.* (2008) identified 18 loci and Croll *et al.* (2008) showed polymorphism in 13 loci, of which two had previously been identified by Raab *et al.* (2005). The two studies used similar, but not identical, strategies to identify repetitive DNA stretches by searching publicly available databases (Table 1). Candidate sequences were then amplified in a set of isolates and potential length polymorphism was scored. In both studies, loci were amplified in a number of isolates from different geographic locations. It should be noted that one locus described by Mathimaran *et al.* (2008) is the same as one polymorphic locus identified by Croll *et al.* (2008) but has been given two different designations. The variation in two more loci reported by Mathimaran *et al.* (2008) is documented in previously published work. We hope that Table 1 will help researchers who intend to use these markers to identify the different loci for which primers have been developed and prevent unintentional studies of the same locus under two different names.

Locus *Glint08* identified by Mathimaran *et al.* (2008) is identical to locus Bg348 from Croll *et al.* (2008), even though the primers are located at different distances from the repetitive sequence region. Loci *Glint09* and *Glint18* identified by Mathimaran *et al.* (2008) were previously published by Corradi & Sanders (2006) and described as genes encoding P-type IID ATPases. Corradi & Sanders (2006) reported

Table 1 Summary of markers developed for *Glomus intraradices*

Locus	Accession no.	Database	Function	Type	Length polymorphism	Reference
Bg32	CG431930	GSS	Unknown	Probably noncoding	Indels	Croll <i>et al.</i> (2008)
Bg42	CG431913	GSS	Unknown	Probably noncoding	(TA) repeat + other indels	Croll <i>et al.</i> (2008)
Bg62	CG431880	GSS	RNA polymerase II large subunit	Proximate coding region	(TAAAA) repeat + other indels	Croll <i>et al.</i> (2008)
Bg196	CG431972	GSS	Unknown	Probably noncoding	Several repeat motifs + other indels	Croll <i>et al.</i> (2008)
Bg235	CG432041	GSS	Unknown	Probably noncoding	Several indels	Croll <i>et al.</i> (2008)
Bg273	CG432137	GSS	Unknown	Probably noncoding	(T) + (A) repeats + other indels	Croll <i>et al.</i> (2008)
Bg276	CG432062	GSS	Unknown	Probably noncoding	Several indels	Croll <i>et al.</i> (2008)
Bg303	CG432175	GSS	Unknown	Probably noncoding	Several indels	Croll <i>et al.</i> (2008)
Bg348	CG432294	GSS	Predicted protein of unknown function	Proximate coding region	(TAA) + (TAAA) repeats + other indels	Croll <i>et al.</i> (2008)
Bg355	CG432269	GSS	Unknown	Probably noncoding	Several indels	Croll <i>et al.</i> (2008)
Nuclear intron	BE603853	EST	Intron in gene of unknown function	Proximate coding region	(T), (A) + (TAA) repeats	Croll <i>et al.</i> (2008)
mtLSU int1	AJ973189-193	Standard	Intron in mitochondrial LSU gene	Proximate coding region	Several indels	Raab <i>et al.</i> (2005); Croll <i>et al.</i> (2008)
mtLSU int2	AJ973189-193	Standard	Intron in mitochondrial LSU gene	Proximate coding region	Indel	Raab <i>et al.</i> (2005); Croll <i>et al.</i> (2008)
<i>Glint01</i>	CG432086+113*	GSS	Unknown	Coding	(AAT) repeat + other indels	Mathimaran <i>et al.</i> (2008)
<i>Glint02</i>	DT883628	EST	Unknown	Coding	(GAA) repeat only?	Mathimaran <i>et al.</i> (2008)
<i>Glint03</i>	BI452162	EST	Unknown	Coding	(TTAT) repeat? + other indels	Mathimaran <i>et al.</i> (2008)
<i>Glint04</i>	BM959176*	EST	Unknown	Coding	(TTA) repeat? + other indels	Mathimaran <i>et al.</i> (2008)
<i>Glint05</i>	BE603957*	EST	Putative cell wall protein	Coding	(TAT) repeat? + other indels	Mathimaran <i>et al.</i> (2008)
<i>Glint06</i>	BM959329	EST	Unknown	Coding	(CAT) repeat? + other indels	Mathimaran <i>et al.</i> (2008)
<i>Glint07</i>	BE603778*	EST	Unknown	Coding	(TTA) repeat? + other indels	Mathimaran <i>et al.</i> (2008)
<i>Glint08</i>	CG432294	GSS	Predicted protein of unknown function	Proximate coding region	(AATA) repeat?	Mathimaran <i>et al.</i> (2008)
(same as Bg348)					but see Bg348 above	
<i>Glint09</i>	AM118108	Standard	P-Type IID ATPase	Coding	(AATG) repeat? + other indels	Corradi & Sanders (2006); Mathimaran <i>et al.</i> (2008)
<i>Glint10</i>	BM027318	EST	Unknown	Coding	(AATGGT) repeat? + other indels	Mathimaran <i>et al.</i> (2008)
<i>Glint11</i>	BI452145	EST	Unknown	Coding	(CAA) repeat only?	Mathimaran <i>et al.</i> (2008)
<i>Glint12</i>	BM959214	EST	Unknown	Coding	(CAA) repeat + other indels	Mathimaran <i>et al.</i> (2008)
<i>Glint13</i>	BM959443*	EST	Unknown	Coding	(AAT) repeat? + other indels	Mathimaran <i>et al.</i> (2008)
<i>Glint14</i>	BM027461*	EST	Unknown	Coding	(T) repeat only?	Mathimaran <i>et al.</i> (2008)
<i>Glint15</i>	BM959581*	EST	Unknown	Coding	(T) repeat only?	Mathimaran <i>et al.</i> (2008)
<i>Glint16</i>	CG431704+705*	GSS	Unknown	Probably noncoding	(A) repeat only?	Mathimaran <i>et al.</i> (2008)
<i>Glint17</i>	CG431789+901*	GSS	Unknown	Probably noncoding	(T) repeat only?	Mathimaran <i>et al.</i> (2008)
<i>Glint18</i>	AM118108	Standard	P-Type IID ATPase	Coding	(A) repeat only?	Corradi & Sanders (2006); Mathimaran <i>et al.</i> (2008)
(same as <i>Glint09</i>)						

Loci are named according to the original publications (Raab *et al.*, 2005; Corradi & Sanders, 2006; Croll *et al.*, 2008; Mathimaran *et al.*, 2008). The putative functions of loci are noted if known from previously published work or if a BLASTX database search on National Center for Biotechnology Information (NCBI) revealed a highly significant match with a known fungal protein (alignment score > 50). Accession numbers show the original sequence of the repeat motif. * denotes accession numbers of loci where highly similar sequences from the database were assembled to make a contig covering the repeat motif. In these cases, the accession number indicates one of the original sequences covering the complete repeat locus. Databases are either the standard nucleotide collection, the genome survey sequences (GSS) or the expressed sequence tag (EST) databases from NCBI. All loci were classified accordingly to their likelihood of being coding or noncoding, depending on whether they are located in an expressed sequence or not. The length polymorphisms among the alleles at each locus were described according to the available sequence data (Croll *et al.*, 2008; Mathimaran *et al.*, 2008); a question mark has been added to the proposed repeat motif if no sequence data were available. For loci where sequence data were not available for all alleles, the length differences among the alleles were used to determine whether the predicted repeat motif alone can explain the observed length polymorphism or whether other indels must be present among the alleles.

polymorphism in a population of *G. intraradices* based on a comparison of different alleles at the same locus. Furthermore, the gene was found to exist in two variants in each of several isolates and in three variants within one isolate (Corradi & Sanders, 2006). Locus *Glint09* is based on the sequence of the third variant; however, the primers designed by Mathimaran *et al.* (2008) are not specific for this particular variant. As a consequence, the primers based on locus *Glint09* potentially amplify up to three different locations in the genome within a single isolate. Locus *Glint18* was identified in an assembled sequence (contig) that matches the P-type IID ATPase variants. However, the resulting consensus sequence does not exactly match any of the original P-type IID ATPase variants, probably as a consequence of the contig being assembled from several different variants (i.e. a chimaeric contig). Consequently, primers for locus *Glint18* do not specifically amplify one of the several variants. Loci *Glint09* and *Glint18* are separated by approx. 500 bp. In our opinion, these two loci are unsuitable for most population genetic studies because of the multi-copy nature of the gene they are located in, unless primer sequences are chosen that restrict the amplification to one variant.

The studies of Mathimaran *et al.* (2008) and Croll *et al.* (2008) both describe polymorphic loci exhibiting size differences of 1 or 2 bp among some alleles. Scoring such a polymorphism is potentially problematic even if PCR products are separated on a capillary sequencer, Spreadex polymer or polyacrylamide gels. These methods offer a high resolution of allele length differences, but the amplification of repeat motifs often leads to the presence of stutter peaks (or shadow bands) as a result of DNA polymerase error. Where small length differences are observed among alleles, it is advisable to obtain sequences that verify that the differences are real and not an artifact of the electrophoresis. This was not done for all loci showing 1- or 2-bp differences in the study by Mathimaran *et al.* (2008) and we suggest more rigorous testing of these differences before using these markers in genotyping studies. If large sets of isolates need to be analysed, the risk of artifacts in the allele identification may be dramatically reduced by using only loci with 3-bp or longer repeat motifs.

Assuming that the length differences are accurate, most of the markers identified by Croll *et al.* (2008) and Mathimaran *et al.* (2008) are useful for demonstrating genetic differences among *G. intraradices* isolates. This does not, however, mean that they are suitable for studying all aspects of AMF population biology. Mutation rates vary across the genome and it is generally assumed that noncoding regions evolve at a higher rate than coding regions, as a result of selective constraints on proteins encoded by the genes. Therefore, it is important to identify the location of the loci in the genome to predict their suitability for particular studies. Mathimaran *et al.* (2008) mostly identified length polymorphism in expressed sequence tags (ESTs). Repeat motifs identified in ESTs are likely to be under selective pressure to maintain functional integrity of the

protein. However, most of the markers reported by Croll *et al.* (2008) and some of those reported by Mathimaran *et al.* (2008) originate from sequences obtained in a genome survey, where regions throughout the genome were randomly sequenced. Because of their random location in the genome, these sequences are likely to be outside of coding regions. However, *G. intraradices* was shown to have a relatively small genome of approx. 15 Mb (Hijri & Sanders, 2004) and, therefore, gene density could be relatively high. Neutral loci are preferable for population genetic studies, as the polymorphism more likely reflects random genetic processes such as mutation, migration or drift. As expected, a majority of the loci from both studies show length polymorphism in the repeat motif. However, a large number of indels and substitutions were also found outside the repeat motif (Table 1). Therefore, the markers do not represent pure simple sequence repeats (or microsatellites) and length differences among alleles should be considered carefully. However, the presence of a large number of substitutions enables researchers to use these markers for a variety of applications such as single nucleotide polymorphism (SNP) genotyping.

Genotyping on a large scale requires amplification of DNA from single spores directly collected from the field, instead of passing through the laborious process of *in vitro* cultivation. However, the small size of *G. intraradices* spores poses a challenge for the amplification of genetic markers because of the very low amount of DNA. Stukenbrock & Rosendahl (2005b) and Mathimaran *et al.* (2008) propose two different approaches to solve this problem. In the first study, a nested PCR was performed and up to five different loci could be amplified. However, it is not known whether this method would perform well with the comparatively small spores of *G. intraradices*. One additional concern is the number of loci that can be amplified simultaneously. Mathimaran *et al.* (2008) chose a promising method called whole-genome amplification (WGA), providing a higher number of template copies of each locus. This method is increasingly used for amplification of DNA from single cells (Spits *et al.*, 2006), unculturable bacteria (Stepanauskas & Sieracki, 2007) or filamentous fungi (Foster & Monahan, 2005), including AMF (Gadkar & Rillig, 2005a,b). While the potential exists to create many template loci from minute samples of cells or spores, several factors may bias the WGA. Notably, WGA is very sensitive to template contamination by other microorganisms as a result of the indiscriminate DNA amplification; a very real concern for spores from pot cultures or the soil (Hijri *et al.*, 2002; Corradi *et al.*, 2004). Furthermore, some parts of the genome tend to be better amplified than others, creating a representation bias in the final product and potentially null alleles (Pinard *et al.*, 2006). In order to apply whole-genome amplification to field-collected spores, the method should be rigorously tested by using well-defined *in vitro* cultivated material as a comparison to whole-genome amplification from single spores of the same culture.

If successfully applied, highly discriminatory markers combined with large-scale hierarchical sampling could elucidate the extent of clonal networks within field sites and resolve patterns of genetic diversity at larger geographic scales. Furthermore, the co-evolution between AMF and their host plants could be studied in detail by identifying spatial distributions of particular genotypes. These areas of investigation have become even more relevant in the context of globally applied inoculum in the absence of data on ecological competitiveness and the potential to persist in the field among native AMF (Schwartz *et al.*, 2006). While the global population genetics of plant pathogenic fungi has received much attention in recent years, studies on plant symbionts will hopefully catch up soon.

Daniel Croll¹, Nicolas Corradi¹, Hannes A. Gamper² and Ian R. Sanders^{1*}

¹Department of Ecology & Evolution, Biophore Building, University of Lausanne, CH-1015 Lausanne, Switzerland;

²Netherlands Institute of Ecology, Center for Terrestrial Ecology, Boterhoeksestraat 48, PO Box 40, NL-6666 ZG, Heteren, The Netherlands (*Author for correspondence: tel +41 21 692 4261; fax +41 21 692 4265; email ian.sanders@unil.ch)

References

- Corradi N, Croll D, Colard A, Kuhn G, Ehinger M, Sanders IR. 2007. Gene copy number polymorphisms in an arbuscular mycorrhizal fungal population. *Applied and Environmental Microbiology* 73: 366–369.
- Corradi N, Kuhn G, Sanders IR. 2004. Monophyly of beta-tubulin and H⁺-ATPase gene variants in *Glomus intraradices*: consequences for molecular evolutionary studies of AM fungal genes. *Fungal Genetics and Biology* 41: 262–273.
- Corradi N, Sanders IR. 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 Evolutionary Biology* 6: 21.
- 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 Phytologist* 178: 672–687.
- Foster SJ, Monahan BJ. 2005. Whole genome amplification from filamentous fungi using Phi29-mediated multiple displacement amplification. *Fungal Genetics and Biology* 42: 367–375.
- Gadkar V, Rillig MC. 2005a. Application of Phi29 DNA polymerase mediated whole genome amplification on single spores of arbuscular mycorrhizal (AM) fungi. *FEMS Microbiology Letters* 242: 65–71.
- Gadkar V, Rillig MC. 2005b. Suitability of genomic DNA synthesized by strand displacement amplification (SDA) for AFLP analysis: genotyping single spores of arbuscular mycorrhizal (AM) fungi. *Journal of Microbiological Methods* 63: 157–164.
- Hijri M, Redecker D, Petetot JA, Voigt K, Wostemeyer J, Sanders IR. 2002. Identification and isolation of two ascomycete fungi from spores of the arbuscular mycorrhizal fungus *Scutellopora castanea*. *Applied and Environmental Microbiology* 68: 4567–4573.
- Hijri M, Sanders IR. 2004. The arbuscular mycorrhizal fungus *Glomus intraradices* is haploid and has a small genome size in the lower limit of eukaryotes. *Fungal Genetics and Biology* 41: 253–261.
- Koch AM, Croll D, Sanders IR. 2006. Genetic variability in a population of arbuscular mycorrhizal fungi causes variation in plant growth. *Ecology Letters* 9: 103–110.
- Mathimaran N, Falquet L, Ineichen K, Picard C, Redecker D, Boller T, Wiemken A. 2008. Microsatellites for disentangling underground networks: strain-specific identification of *Glomus intraradices*, an arbuscular mycorrhizal fungus. *Fungal Genetics and Biology* 45: 812–817.
- Munkvold L, Kjoller R, Vestberg M, Rosendahl S, Jakobsen I. 2004. High functional diversity within species of arbuscular mycorrhizal fungi. *New Phytologist* 164: 357–364.
- Öpik M, Moora M, Liira J, Zobel M. 2006. Composition of root-colonizing arbuscular mycorrhizal fungal communities in different ecosystems around the globe. *Journal of Ecology* 94: 778–790.
- Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, Leamon JH. 2006. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* 7: 216.
- Raab PA, Brennwald A, Redecker D. 2005. Mitochondrial large ribosomal subunit sequences are homogeneous within isolates of *Glomus* (arbuscular mycorrhizal fungi, Glomeromycota). *Mycological Research* 109: 1315–1322.
- Rosendahl S. 2008. Communities, populations and individuals of arbuscular mycorrhizal fungi. *New Phytologist* 178: 253–266.
- Sanders IR, Alt M, Gropp 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 Phytologist* 130: 419–427.
- Schwartz MW, Hoeksema JD, Gehring CA, Johnson NC, Klironomos JN, Abbott LK, Pringle A. 2006. The promise and the potential consequences of the global transport of mycorrhizal fungal inoculum. *Ecology Letters* 9: 501–515.
- Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K. 2006. Whole-genome multiple displacement amplification from single cells. *Nature Protocols* 1: 1965–1970.
- Stepanaukas R, Sieracki ME. 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proceedings of the National Academy of Sciences, USA* 104: 9052–9057.
- Stukenbrock EH, Rosendahl S. 2005a. Clonal diversity and population genetic structure of arbuscular mycorrhizal fungi (*Glomus* spp.) studied by multilocus genotyping of single spores. *Molecular Ecology* 14: 743–752.
- Stukenbrock EH, Rosendahl S. 2005b. Development and amplification of multiple co-dominant genetic markers from single spores of arbuscular mycorrhizal fungi by nested multiplex PCR. *Fungal Genetics and Biology* 42: 73–80.
- Van der Heijden MGA, Klironomos JN, Ursic M, Moutoglou 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.

Key words: arbuscular mycorrhizal fungi (AMF), *Glomus intraradices*, microsatellites, multilocus genotyping, population genetics, simple sequence repeats (SSR).