

Determining a minimum detection threshold in terminal restriction fragment length polymorphism analysis

Kevin C. Courtney*, Luke D. Bainard*, Benjamin A. Sikes \diamond , Alexander M. Koch \S ,
Miranda M. Hart \S , Hafiz Maherali*, John N. Klironomos \S

*Department of Integrative Biology, University of Guelph, 50 Stone Road East, Guelph, ON, N1G 2W1, Canada

\diamond Section of Integrative Biology, The University of Texas, 1 University Station, Austin, TX 78712, USA

\S Biology and Physical Geography Unit, The University of British Columbia – Okanagan, 3333 University Way, Kelowna, BC, V1V 1V7, Canada

Abstract

Terminal restriction fragment length polymorphism (T-RFLP) analysis is a common technique used to characterize soil microbial diversity. The fidelity of this technique in accurately reporting diversity has not been thoroughly evaluated. Here we determine if rare fungal species can be reliably detected by T-RFLP analysis. Spores from three arbuscular mycorrhizal fungal species were each mixed at a range of concentrations (1%, 10%, 50%, and 100%) with *Glomus irregulare* to establish a minimum detection threshold. T-RFLP analysis was capable of detecting diagnostic peaks of rare taxa at concentrations as low as 1%. The relative proportion of the target taxa in the sample and DNA concentration influenced peak detection reliability. However, low concentrations produced small, inconsistent electropherogram peaks contributing to difficulty in differentiating true peaks from signal noise. The results of this experiment suggest T-RFLP is a reproducible and high fidelity procedure, which requires careful data interpretation in order to accurately characterize sample diversity.

Keywords: T-RFLP, detection threshold, arbuscular mycorrhizal fungi, spores

Abbreviations: AM, Arbuscular mycorrhizal; ANOVA, Analysis of Variance; INVAM, International Culture Collection of (Vesicular) Arbuscular Mycorrhizal Fungi; LSU, Large ribosomal subunit; PCR, Polymerase Chain Reaction; P/E, Primer/Enzyme; T-RFLP, Terminal restriction fragment length polymorphism.

1. Introduction

Microbial community composition is highly variable and diverse. Estimated numbers of microbial species far exceed current numbers of described species (Bull &

Stach 2004, Hawksworth 2001). Greater than ninety percent of microorganisms cannot be grown in culture. This inability to maintain the majority of microbes in axenic culture exemplifies the need for culture independent approaches in microbial characterization (Marsh 1999, Nocker 2007, Thies 2007). At the same time, these undescribed microbes represent the most likely sources of new novel compounds (Tan 2006). Therefore it is critical to develop and optimize detection methods that accurately quantify microbial diversity and identify novel taxa. The use of molecular techniques has allowed for increased precision in identification of organisms in natural and artificial communities (Ranjard *et al.* 2000). However, these techniques can also be variable depending on the composition of communities that are sampled (Lekberg *et al.* 2007).

Terminal restriction fragment length polymorphism (T-RFLP) is a genetic fingerprinting technique that has been widely used in microbial research for species identification and diversity characterizations (Thies 2007). This technique has been applied to characterize archaeal, bacterial and fungal communities (Donovan 2004, Aslanian 2006, Mummey & Rillig 2008). T-RFLP allows for high throughput sample processing, which facilitates analysis of complex communities (Tiedje 1999). The technique combines polymerase chain reaction (PCR) with restriction digestion to cleave amplified DNA at specific target sites and generate predictable and reproducible fragments. The distinct cleavage patterns of different taxa can then be compiled into a reference database of diagnostic fragment sizes. Database T-RFLP is used to identify individual taxa within samples by comparing known species-specific electropherogram peaks against unknown samples for identification (Avis *et al.* 2006, Dickie & FitzJohn 2007).

Although T-RFLP is widely used to assess microbial community composition, taxa present in low concentrations may not be detected by T-RFLP analysis (Dickie *et al.* 2002). A taxon is known to be present if one can identify a diagnostic fragment, yet the absence of a diagnostic fragment may represent a detection failure based on the rarity of the target taxon within a sample. If rare taxa are overlooked by T-RFLP, the technique may underestimate diversity and confound our ability to distinguish between samples that differ only in species relative abundance rather than presence/absence.

The recent proliferation of next generation sequencing has led some speculation that T-RFLP is no longer a powerful approach for studying microbial communities because of it is unable to reliably characterize species in low abundance (Hartmann and Widmer 2008, Hadamay and Knight 2009, Fierer and Lennon 2011). This is a real problem for microbial ecologists, since the vast majority of microbial communities are dominated by rare species (Sogin *et al.* 2006, Fierer and Lennon 2011, Reid and Buckley 2011).

In this study, we tested the limits of T-RFLP to detect microorganisms in low abundance, using artificial communities. Arbuscular mycorrhizal (AM) fungi were used as the model system because of their relatively large propagules, which facilitated the isolation of single spores. AM fungal species were selected from distinct phylogenetic lineages and diluted in a sample of *Glomus irregulare* spores at 1, 10, 50 and 100% concentrations. *G. irregulare* (DAOM 197198) was chosen as the background fungus due to its wide global distribution and common use as a model fungus in molecular studies (Blaszkowski *et al.* 2009, Stockinger *et al.* 2009, Sokolski *et al.* 2010). We tested

if the presence or absence of taxon specific electropherogram peaks, was influenced by the relative abundance of the taxa in a sample. In addition to relative abundance, we investigated whether spore size and DNA content influenced species detection by T-RFLP. We postulate that T-RFLP efficacy is dictated by species relative abundance, which may ultimately be driven by DNA content.

2. Materials and Methods

2.1 Spore extraction

AM fungal inoculum was provided by the International Culture Collection of (Vesicular) Arbuscular Mycorrhizal Fungi (INVAM) and propagated on a leek (*Allium porrum*) host. *Glomus irregulare* was designated as the background fungus to be present in all mixed samples. *Glomus clarum* (NB112A), *Glomus etunicatum* (MX116A), and *Gigaspora albida* (CL151) were selected to represent rare or underrepresented taxa in mycorrhizal communities. The rare taxa were chosen to be increasingly different phylogenetically from *G. irregulare*. In addition, the spores of these fungi vary in size, ranging from 93 μm to 250 μm (Table 1). *G. irregulare* spores were provided by Premier Tech Ltd. (Rivière-du-Loup, Quebec, Canada). Rare taxa spores were isolated by wet sieving and sucrose centrifugation from the INVAM propagated cultures. Spores of average size and color were selected from the centrifugation product by pipette, and placed into microcentrifuge tubes. Each sample was replicated 5 times and contained 100 spores in total, with increasing proportions of rare spores. The three rare taxa were individually mixed at the following ratios with *G. irregulare*: 1:99, 10:90, 50:50, and 100:0 spores.

2.2 DNA Extraction

Each 100-spore sample was sonicated (Branson 2510 ultrasonic cleaner) for one minute, rinsed in sterile water and placed in 1.5 ml microcentrifuge tubes. Spores were crushed in 40 μl of 1x TE buffer by micropestle and boiled for two minutes. Samples were centrifuged at 13,000 rpm for one minute in the presence of 10 μl of 20% Chelex 100 (BioRad) to pellet cell wall fragments and 20 μl of supernatant was transferred to new tubes (Van Tuinen *et al.* 1998). The extracts were subjected to an ethanol precipitation to purify DNA and resuspended in 1x TE buffer for downstream applications. DNA concentrations from each sample were determined using a Nanodrop 8000 (Nanodrop Technologies) spectrophotometer.

2.3 PCR Amplification

The 28S large ribosomal subunit (LSU) gene was targeted using the FLR3/FLR4 primer pair, which specifically amplifies AM fungi (Gollote *et al.* 2004) and has been used effectively to identify AM fungi in mixed communities (Mummey and Rillig 2007). Each 25 μl reaction contained 1 μl of DNA extract, 2.5 μl PCR buffer, 1.5 mM MgCl_2 , 200 μM dNTPs, 0.5 U Taq, and 0.5 μM HEX labeled FLR3 and FAM labeled FLR4 primers. Thermocycling (BioRad) conditions consisted of an initial denaturing step of 95°C for 3 minutes, followed by 35 cycles of 93°C for 1 minute, 58°C 1 minute and 72°C for 1 minute, and a final extension period of 72°C for 10 minutes. To account for potential methodological error, the PCR protocol was replicated for each sample to generate a duplicate T-RFLP dataset. Successful amplification was verified on a 1.5% agarose gel.

PCR products were then purified using the Qiagen QIAquick PCR purification kit (Qiagen) and eluted in 30µl of elution buffer.

2.4 Digestion

Purified PCR products were digested with the restriction enzymes *AluI* and *MboI* (Invitrogen) in separate reactions. Digestion reactions consisted of 5 µl of cleaned PCR product and 1 U *AluI* or *MboI* in the manufacturers recommended buffer. Samples were incubated for 6 hours at 37°C followed by 20 minutes at 65°C to inactivate the enzymes. As a result of amplifying the template DNA with fluorescent-labeled forward and reverse primers and digesting each sample with two enzymes, each sample produced four terminal restriction fragment datasets. Analysis of the resultant T-RFLP fragments was performed using an Applied Biosystems 3730 DNA sequencer and Genemapper v3.4 software (Applied Biosystems) to include fragments that ranged from 40 to 400 base pairs in length.

2.5 Data Analysis

Raw Genemapper output was manipulated with the Excel based macro, Treeflap, (<http://www.sci.monash.edu.au/wsc/staff/walsh/treeflap.xls>) to produce a binary presence/absence matrix for all sample peaks, and to reorganize the non-binary data into a suitable format for import into the statistical analysis program R (R Development Core Team 2008). A Jaccardian dissimilarity matrix was constructed from binary data using Vegan (1.15-3), an R based community ecology analysis package, to determine if sample replicates produced similar peak profiles (Oksanen *et al.* 2008). An analysis of variance (ANOVA) was conducted on the dissimilarity data and pair wise peak profile comparisons were made by post-hoc Tukey HSD testing using R. The single taxa samples (non-binary) were sorted and aligned within Excel and examined to identify diagnostic species peaks for each primer/enzyme (P/E) combination (e.g. four diagnostic peaks per species). A conservative method of diagnostic peak selection was used by selecting fragments that were present in all samples containing only the target species (100%), and absent from all samples lacking the target species. Each P/E combination per taxa contained at least one unique peak that differed from all other peaks in the dataset by ± 0.5 bases. However, *G. clarum* only had a unique restriction fragment size in 3 out of 4 P/E combinations. Mixed samples were then searched manually for diagnostic peaks to confirm the presence or absence of each fungal species. Two-way ANOVA and Tukey HSD tests were done in R to investigate the effects of taxa and spore abundance on diagnostic peak detection rates with a significance level of $\alpha=0.05$.

3. Results

3.1 Binary dissimilarity

Replicate DNA extractions and T-RFLP procedures produced similar peak profiles. Replicated samples were consistently more similar to themselves than all other samples. One hundred, fifty, ten and single spore samples had a dissimilarity mean and standard deviation of 0.51 (0.16), 0.79 (0.12), 0.85 (0.06) and 0.93 (0.04) respectively. With the exception of *G. clarum*'s 10-spore/50-spore comparison ($p=0.01$); significant differences in peak profiles were observed between each treatment ($p<0.0001$). As target spore

number increased, the peak profiles became more similar to their respective single taxa peak profiles (Fig. 1). Conversely, mixed samples became more similar to the *G. irregulare* profile as rare spore number decreased.

3.2 Spore number

T-RFLP was capable of detecting diagnostic peaks from single spores in mixed samples for all treatment replicates in at least one out of four P/E combinations. However, multiple P/E combinations are commonly used in practice to rule out false positive detection (Dickie and FitzJohn 2007). As we increased the stringency of our detection criteria and required more P/E combinations to confirm taxon presence percent detection decreased (Fig. 2). As expected, increasing spore number improved peak detection and increased the number of P/E combinations detected per sample. Small diagnostic peaks (values <500 relative fluorescence units) were associated with low spore numbers and the largest peaks were generated from 50 or 100 spore samples.

The effect of spore number on detection rate was significantly different between treatments regardless of taxa ($p < 0.0001$). Both the number of target spores and taxa determined detection rate, however little interaction was observed between independent variables ($p=0.14$). The greatest differences in spore numbers were between the single spore and multi-spore samples (Fig. 2). Differences in detection became less statistically significant as spore number increased.

3.3 Spore size

Spore size was also correlated with detection rate, which is likely due to differences in DNA content (Table 1). The three target taxa had very different spore sizes. Results indicate DNA concentration was correlated with average spore size ($p=0.0002$). We found target taxa were better detected as spore size increased. *Gi. albida* produces the largest spores and was detected most effectively; particularly with low spore numbers (Fig. 2).

4. Discussion

Our results indicate that T-RFLP can reproducibly identify rare taxa present at low concentrations in simple artificial communities. Species abundance clearly had an effect on detection; rare taxa were consistently detected in 10% samples with a significant decline in 1% samples. These results indicate that there is likely a threshold between 1-10% for rare taxa in the community, above which target taxa are more likely to be detected using multiple primer/enzyme combinations with T-RFLP.

These experiments likely represent a highly simplified system from field settings where T-RFLP is thought to be most useful; therefore detection thresholds may be even higher in practice. AM fungal communities in the field contain more complex communities than only two taxa (Maherali and Klironomos 2007). In addition, PCR of any target microbial group also encounters DNA from non-target organisms such as other fungi, protists, bacteria and plants. Finally, if microbial communities obey species abundance

distributions (SAD) (Moyer *et al.* 1994, Martiny *et al.* 2006) many species in a sample may be below the 1-10% threshold we report here.

Management of T-RFLP data also influences results because community diversity can be exaggerated when false peaks are not discarded or underestimated by strict minimum peak height thresholds (Avis *et al.* 2006, Dickie *et al.* 2002). By retaining all peak heights in this experiment, we found small but reproducible diagnostic peaks of rare taxa can fall within this range of false peaks and may be cut off by thresholds.

In addition to spore frequency, the size and number of spores present in a mixed sample influences the likelihood of detection. The samples with large spores or higher numbers of spores contained more DNA; as DNA concentration increases, detection rate improved. AM fungi exhibit a wide range of spore sizes and reproduction rates and their specific life strategies will influence detection by T-RFLP (Hart and Reader 2002). Taxa that are rare and produce large spores may be detected as effectively as taxa that are more abundant but generate smaller spores.

These results show taxa present at concentrations as low as 1% in a sample can generate small electropherogram peaks by T-RFLP. Field and greenhouse experiments, however, will not generate such clean, reproducible and distinguishable peak profiles (Avis *et al.* 2006). When working with soil derived DNA, small peaks can be eliminated by the creation of a minimum detection threshold, which is often applied to clean up the data (Abdo *et al.* 2006, Dunbar *et al.* 2001, Osborne *et al.* 2006). Thresholds are a logical method for eliminating small noisy peaks, while enhancing relevant peaks. However, small peaks may represent true taxon-specific markers, which could be lost in data manipulation procedures. The high fidelity of T-RFLP found in this study suggests that data manipulation results in an underrepresentation of diversity when peak height thresholds are applied to raw data. Small peaks are indistinguishable from noise and in many cases this represents inherent error associated with T-RFLP analysis. This finding supports the utility of database T-RFLP, where taxa within a community have known diagnostic peaks that can be separated from noise, rather than removing all noise and rare taxa together.

Conclusion

In evaluating T-RFLP, AM fungi are a useful study system. Unlike many other microorganisms, large AM fungal spores can be easily manipulated by pipette, facilitating single spore isolation. However, the utility of T-RFLP for describing microbial community structure is still debated because of its inability to detect low frequency sequences. Here we showed that in artificially constructed systems, one AM fungal spore can be detected by T-RFLP in a mixed sample. This suggests that this technique has the potential to reliably identify taxa present at low frequencies, at least for AMF fungi, or in systems where that are not dominated by rare taxa, such as industrial sites, or extreme environments. Given that all current approaches, including high throughput sequencing, have difficulties resolving very rare taxa in

mixed assemblages, T-RFLP is still a robust and very affordable approach for specific questions that don't require absolute quantification of taxa, or those looking at systems that have more balanced species abundance.

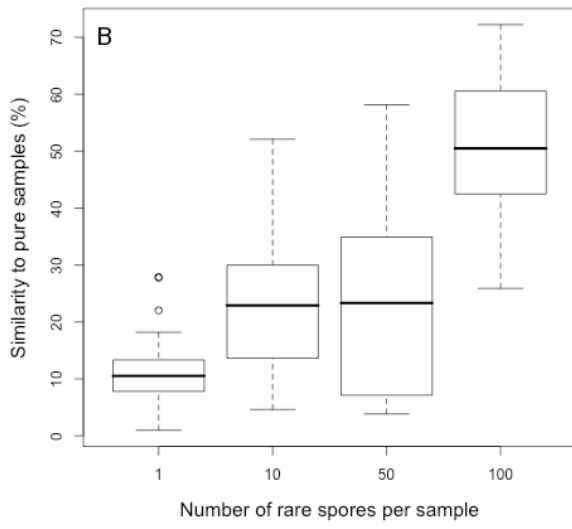
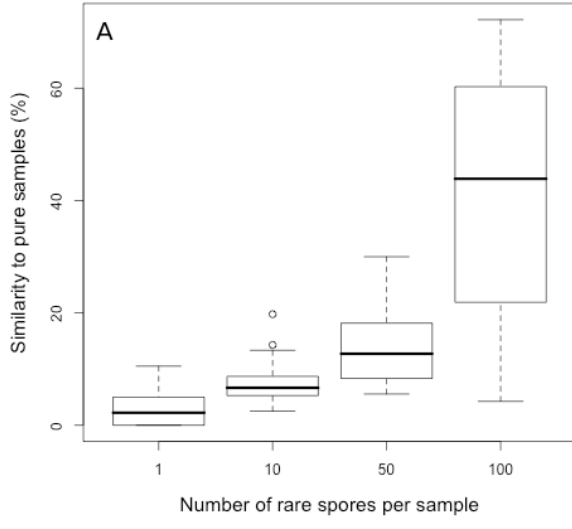
References

- Abdo, Z., Schütte, U.M.E., Bent, S.J., Williams, C.J., Forney, L.J., Joyce, P. (2006) Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes. *Environmental Microbiology* **8**, 929–938.
- Avis, P.G., Dickie, I.A., Mueller, G.M. (2006) A 'dirty' business: testing the limitations of terminal restriction fragment length polymorphism (T-RFLP) analysis of soil fungi. *Molecular Ecology* **15**, 873-882.
- Blaszkowski, J., Czerniawska, B., Wubet, T., Schaefer, T., Buscot, F., Renker, C. (2009) *Glomus irregulare*, a new arbuscular mycorrhizal fungus in the Glomeromycota. *Mycotaxon* **106**, 247-267.
- Bull, A.T., Stach, J.E.M. 2004. An overview of biodiversity - estimating the scale. In: Bull, A.T. (Eds.), *Microbial Diversity and Bioprospecting*. ASM press, Washington, pp. 15-28.
- Dickie, I.A., FitzJohn, R.G. (2007) Using terminal restriction fragment length polymorphism (T-RFLP) to identify mycorrhizal fungi: a methods review. *Mycorrhiza* **17**, 259–270.
- Dickie, I.A., Xu, B., Koide, R.T. (2002) Vertical niche differentiation of ectomycorrhizal hyphae in soil as shown by T-RFLP analysis. *New Phytologist* **156**, 527-535.
- Donovan, S.E., Purdy, K.J., Kane, M.D., Eggleton, P. (2004) Comparison of *Euryarchaea* Strains in the Guts and Food-Soil of the Soil-Feeding Termite *Cubitermes fungifaber* across Different Soil Types. *Applied and Environmental Microbiology* **70**, 3884-3892.
- Dunbar, J., Ticknor, L.O., Kuske, C.R. (2001) Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. *Applied and Environmental Microbiology* **67**, 190–197.
- Fierer, N., Lennon, J.T. (2011) "The generation and maintenance of diversity in microbial communities. *American Journal of Botany* **98**: 439-448.
- Gollotte, A., van Tuinen, D., Atkinson, D. (2004) Diversity of arbuscular mycorrhizal fungi colonizing roots of the grass species *Agrostis capillaries* and *Lolium perenne* in a field experiment. *Mycorrhiza* **14**, 111-117.

- Hamady, M., Knight, R. (2009) Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research* **19**: 1141-1152.
- Hart, M.M., Reader, R.J. (2002) Taxonomic basis for variation in the colonization strategy of arbuscular mycorrhizal fungi. *New Phytologist* **153**, 335-344.
- Hartmann, M., Widmer, F. (2008) Reliability for detecting composition and changes of microbial communities by T-RFLP genetic profiling. *FEMS Microbial Ecology* **63**: 249-260.
- Hawksworth, D.L. (2001) The magnitude of fungal diversity: the 1.5 million species revised. *Mycological Research* **105**, 1433-1432.
- Lekberg, Y., Koide, R.T., Rohr, J.R., Aldrich-Wolfe, L., Morton J.B., (2007) Role of niche restrictions and dispersal in the composition of arbuscular mycorrhizal fungal communities. *Journal of Ecology* **95**, 95–105.
- Liu, W.T., Marsh, T.L., Cheng, H., Forney, L.J. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied Environmental Microbiology* **63**, 4516–4522.
- Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology* **4**: 102-112.
- Moyer, C.L., Dobbs, F.C., Karl, D.M. (1994) Estimation of Diversity and Community Structure through Restriction Fragment Length Polymorphism Distribution Analysis of Bacterial 16S rRNA Genes from a Microbial Mat at an Active, Hydrothermal Vent System, Loihi Seamount, Hawaii. *Applied and Environmental Microbiology* **60**, 871-879.
- Mummey, D.L., Rillig, M.C. (2008) Spatial characterization of arbuscular mycorrhizal fungal molecular diversity at the submetre scale in a temperate grassland. *FEMS Microbiology Ecology* **64**, 260-270.
- Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. (2008) vegan: Community Ecology Package. R package version 1.15-1. <http://cran.r-project.org/>, <http://vegan.r-forge.r-project.org/>
- Osborne, C.A., Rees, G.N., Bernstein, Y., Janssen, P.H. (2006) New threshold and confidence estimates for terminal restriction fragment length polymorphism analysis of complex bacterial communities. *Applied Environmental Microbiology* **72**, 1270–1278.
- Ranjard, L., Poly, F., Nazaret, S. (2000) Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. *Research in Microbiology* **151**, 167-177.
- Reid, A., Buckley, M. (2011) The Rare Biosphere. A Report from the American

Academy of Microbiology, Washington DC.

- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, M.D., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America* **103**: 12115-12120.
- Sokolski, S., Dalpe, Y., Seguin, S., Khasa, D., Levesque, C.A., Piche, Y. (2010) Conspicuity of DAOM 197198, the model arbuscular mycorrhizal fungus, with *Glomus irregulare*: molecular evidence with three protein-encoding genes. *Botany* **88**, 829-838.
- Stockinger, H., Walker, C., Schüßler, A. (2009) 'Glomus intraradices DAOM 197198', a model fungus in arbuscular mycorrhizal research, is not *Glomus intraradices*. *New Phytologist* **183**, 1176-1187.
- Tan, F., Gyllenhaal, C., Soejarto, D.D. (2006) Biodiversity as a source of anticancer drugs. *Current Drug Targets* **7**, 265-277.
- Van Tuinen, D., Jacquot, E., Zhao, B., Gollotte, A., Gianinazzi-Pearson, V. (1998) Characterization of root colonization profiles by a microcosm community of arbuscular mycorrhizal fungi using 25S rDNA-targeted nested PCR. *Molecular Ecology* **7**, 879-887.



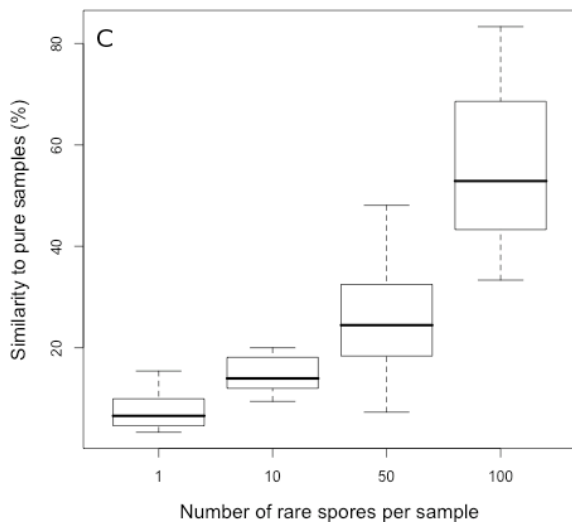


Fig 1 – Jaccardian dissimilarity plot - **A** - *Gi. albida*, **B** - *G. clarum*, **C** - *G. etunicatum*. Binary T-RFLP profiles were compared between mixed and single-species samples. Jaccardian dissimilarity matrix was used to calculate similarity between profiles. Mixed sample T-RFLP profiles become more similar to single-species profiles as the number of target spores increases.

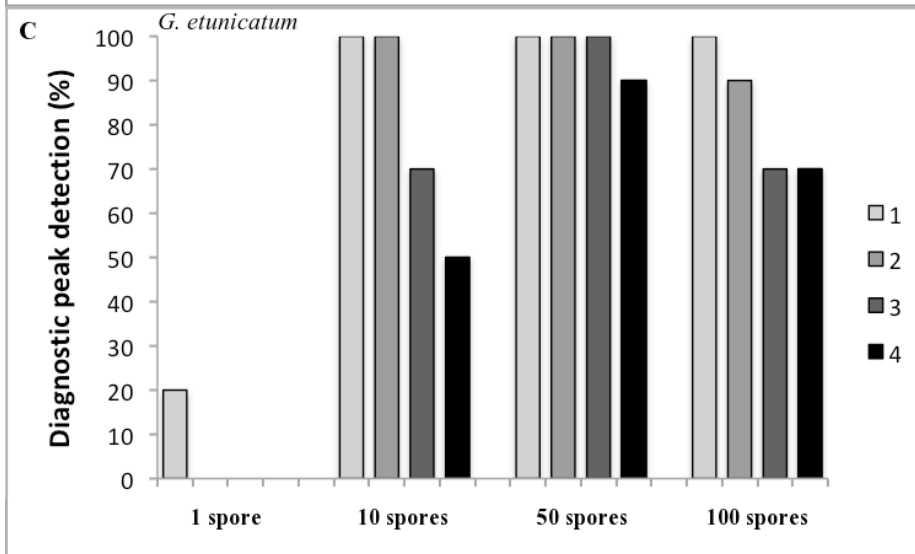
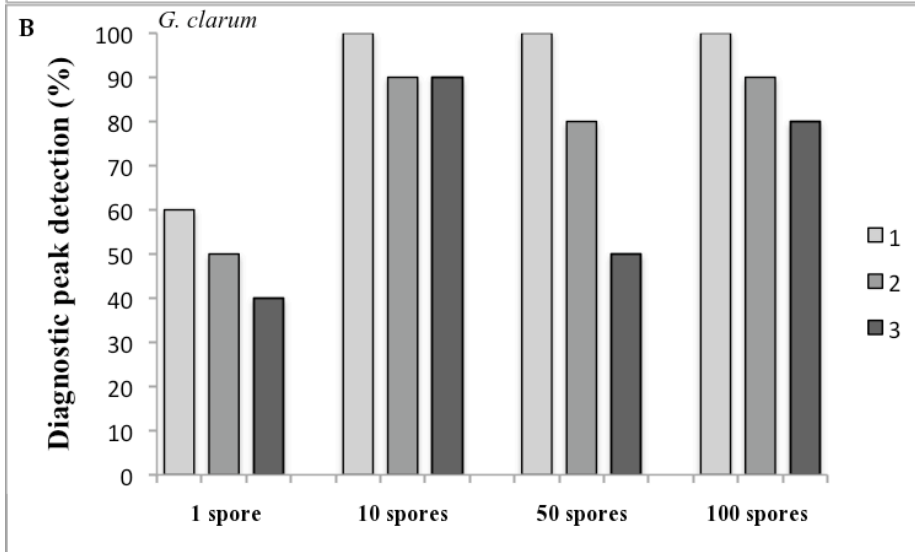
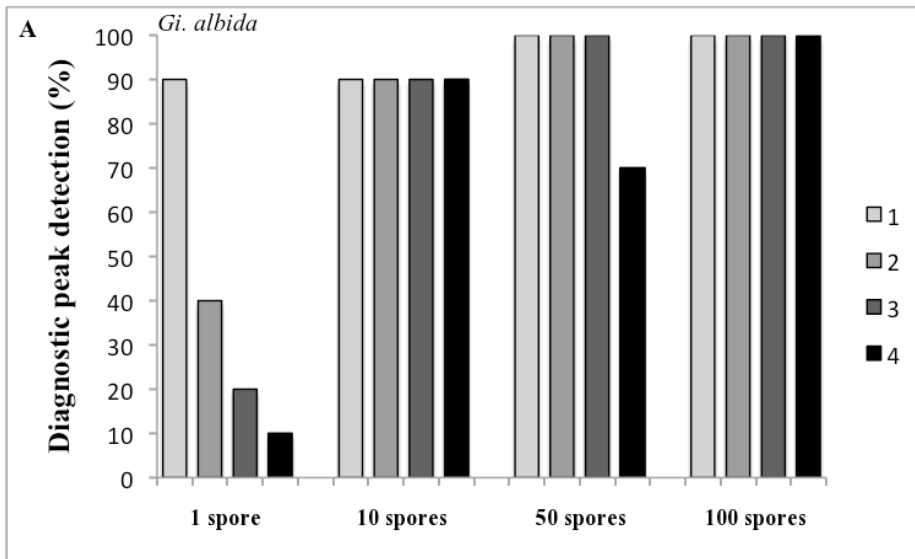


Fig. 2 - Diagnostic peak detection - **A** – *Gi. albida*, **B** – *G. clarum*, **C** – *G. etunicatum*. Presence/ absence frequencies of detecting diagnostic peaks calculated based on the number of spores present in mixed samples. Coloured bars represent detection when 1 to 4 primer/enzyme combinations are used, showing detection declines with greater numbers of primer/enzyme combinations.

Species	DNA concentration (ng ul ⁻¹)	Spore diameter (μm)
<i>Gigaspora albida</i>	17.938 (1.51)	250
<i>Glomus clarum</i>	3.724 (0.28)	182
<i>Glomus etunicatum</i>	0.311 (0.49)	129
<i>Glomus irregulare</i>	NA*	93

Table 1 - Summary of mean DNA concentrations (with standard error) measured from 5 replicate single-species extractions containing 100 spores and mean spore diameters (INVAM). **G. irregulare* concentrations were below the Nanodrop-8000 detection limit.