



# What Goes in Must Come out: Testing for Biases in Molecular Analysis of Arbuscular Mycorrhizal Fungal Communities

T. E. Anne Cotton<sup>1,2,\*</sup>, Alex J. Dumbrell<sup>2</sup>, Thorunn Helgason<sup>1</sup>

<sup>1</sup> Department of Biology, University of York, York, United Kingdom, <sup>2</sup> School of Biological Sciences, University of Essex, Colchester, United Kingdom

## Abstract

Arbuscular mycorrhizal (AM) fungi are widely distributed microbes that form obligate symbioses with the majority of terrestrial plants, altering nutrient transfers between soils and plants, thereby profoundly affecting plant growth and ecosystem properties. Molecular methods are commonly used in the study of AM fungal communities. However, the biases associated with PCR amplification of these organisms and their ability to be utilized quantitatively has never been fully tested. We used Terminal Restriction Fragment Length Polymorphism (TRFLP) analysis to characterise artificial community templates containing known quantities of defined AM fungal genotypes. This was compared to a parallel *in silico* analysis that predicted the results of this experiment in the absence of bias. The data suggest that when used quantitatively the TRFLP protocol tested is a powerful, repeatable method for AM fungal community analysis. However, we suggest some limitations to its use for population-level analyses. We found no evidence of PCR bias, supporting the quantitative use of other PCR-based methods for the study of AM fungi such as next generation amplicon sequencing. This finding greatly improves our confidence in methods that quantitatively examine AM fungal communities, providing a greater understanding of the ecology of these important fungi.

**Citation:** Cotton TEA, Dumbrell AJ, Helgason T (2014) What Goes in Must Come out: Testing for Biases in Molecular Analysis of Arbuscular Mycorrhizal Fungal Communities. PLoS ONE 9(10): e109234. doi:10.1371/journal.pone.0109234

**Editor:** Raffaella Balestrini, Institute for Sustainable Plant Protection, C.N.R., Italy

**Received:** May 3, 2014; **Accepted:** September 1, 2014; **Published:** October 2, 2014

**Copyright:** © 2014 Cotton et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All sequence files are available from the EMBL database (accession numbers KJ749695-KJ749700).

**Funding:** This study was funded by a Biotechnology and Biological Sciences Research Council (www.bbsrc.ac.uk) Ph.D. studentship (TEAC) and the Natural Environment Research Council (www.nerc.ac.uk) grant NE/D01090X/1 (AJD, TH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: anne.cotton@sheffield.ac.uk

‡ Current address: Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom

## Introduction

Arbuscular mycorrhizal (AM) fungi form the most abundant symbiosis on earth, colonising the roots of approximately two thirds of terrestrial plants [1,2]. In addition to being extremely common, they are also highly influential in ecosystems, affecting plant growth and communities [3–6], driving nutrient cycles [7,8], determining soil properties [8,9] and interacting with many other soil organisms [10,11]. However, AM fungi are extremely difficult to study by non-molecular approaches due to their cryptic morphology and complex growth requirements. Thus, molecular methods have revolutionised our understanding of these organisms and are amongst the most promising tools for their study [12,13].

One of the most widely used DNA-based methods is the analysis of Terminal Restriction Fragment Length Polymorphisms (TRFLPs; e.g. refs [14–17]). TRFLP analysis involves polymerase chain reaction (PCR) amplification of environmental DNA extracts using fluorescently labelled primers, restriction digestion of the resulting terminally labelled amplicons and measurement of the sizes of the terminal restriction fragments (TRFs) produced [18]. This produces community ‘fingerprints’ for samples analysed that can be compared in terms of diversity and composition. The use of TRFLPs alone does not produce sequences of the organisms

in a community, but if used in conjunction with sequence databases, putative identities may be assigned to TRFs, allowing their identification [13,19].

Many studies have used TRFLPs and other molecular methods such as next generation sequencing (NGS) to study AM fungal communities (e.g. refs [14–17,20,21]). Most of these studies have made the major assumption that there are minimal biases in these molecular methods which could influence the quantification of AM fungal communities. Yet these methods all involve PCR, and PCR bias, whereby different genotypes are preferentially amplified due to their nucleotide composition or amplification reaction kinetics (reviewed in ref. [22]), is therefore possible. It has been argued that PCR amplification of AM fungi with the commonly used primers AM1 and NS31 is unlikely to be biased due to the conserved nature of their primer binding regions and low variability in GC content in the resulting amplicons [14], but this has never been empirically tested. If such biases do exist, the results from such studies could only be used qualitatively, preventing more informative, quantitative analyses. However, if such biases are not present, it would allow the differentiation of communities that contain the same components, but in different

relative abundances and the calculation of diversity estimates based on evenness as well as richness.

In this study, we created artificial community templates containing known ratios of characterised AM fungal genotypes and analysed them using TRFLP analysis. The relative abundances of TRFs that were generated and the modelled relative abundances produced assuming no bias in the PCR and TRFLP process, were analysed and compared. This experiment thereby tested the hypothesis that TRFLP analysis accurately measures the relative differences between different AM fungal communities in terms of their composition, diversity and structure.

**Materials and Methods**

**Generating community template components**

Artificial community templates were produced by mixing PCR products from 18S rDNA clone libraries produced for a study of AM fungi in soybean roots [23]. In summary, total community DNA was extracted from soybean roots using a PowerPlant DNA isolation Kit (Mo Bio Laboratories, Inc., CA, USA) according to the manufacturer’s instructions. Partial small subunit (SSU) ribosomal DNA fragments (ca. 550 bp) were amplified with the universal eukaryotic primer NS31 [24] and the AM fungal specific primer AM1 [25]. PCR products were cleaned using a QIAquick PCR Purification Kit (Qiagen Ltd, West Sussex, UK) and cloned into pGEM-T Easy Vector (Promega, WI, USA) by incubating 3µl of cleaned PCR product with 5µl of ligation buffer, 1µl T4 DNA ligase and 1µl of pGEM-T Easy Vector for 3 hours at room temperature. Vectors (4µl) were transformed into *Escherichia coli* cells by incubation with 25µl competent *E. coli* cells (DH5α; Invitrogen, Renfrewshire, UK) at 4°C for 30 minutes, followed by a heatshock of 42°C for 45 seconds and rotary incubation with 475µl Super Optimal broth with Catabolite repression (SOC) at 37°C for 1 hour. The resulting bacteria were plated onto LB-agar amended with X-gal and ampicillin. These were incubated at 37°C overnight before storing at 4°C. Putative positive transformants were amplified using primers SP6 and T7 (as described in ref. [26]), which bind to the plasmid DNA on either side of the DNA insert, producing ca. 725 bp amplicons containing the 18S AM fungal insert flanked by DNA derived from the plasmid. These were sequenced (Technology Facility, University of York, UK) using primer SP6 and BigDye™ terminator cycling conditions on an ABI 3130 capillary sequencer (Applied Biosystems Inc, Applied Biosystems Inc., CA, USA). Sequences were aligned using ClustalX [27] with default parameter settings and pairwise sequence similarities were calculated. Six distinct sequence types (genotypes) were chosen to generate the artificial DNA templates created, which when combined met three criteria: 1) virtual digestion *in silico* suggested they would each produce uniquely sized TRFs with each of two enzymes commonly used in TRFLP analyses: *Alu-I* and *Hinf-I*; 2) the AM1/NS31 insert had the same orientation within the plasmid; 3) each genotype showed ≤97% pairwise sequence similarity to all other genotypes used, reflecting commonly used cut-offs for defining Operational Taxonomic Units (OTUs). These sequences have been submitted to the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database (accession numbers KJ749695-KJ749700).

To generate sufficient genetic material to produce the templates, PCR products derived from the chosen genotypes were amplified again using the same protocol as before (with primers SP6 and T7), but with only 15 cycles to reduce the likelihood of errors. The resulting PCR products were purified using a QIAquick PCR Purification Kit (Qiagen) and divided into three separate aliquots

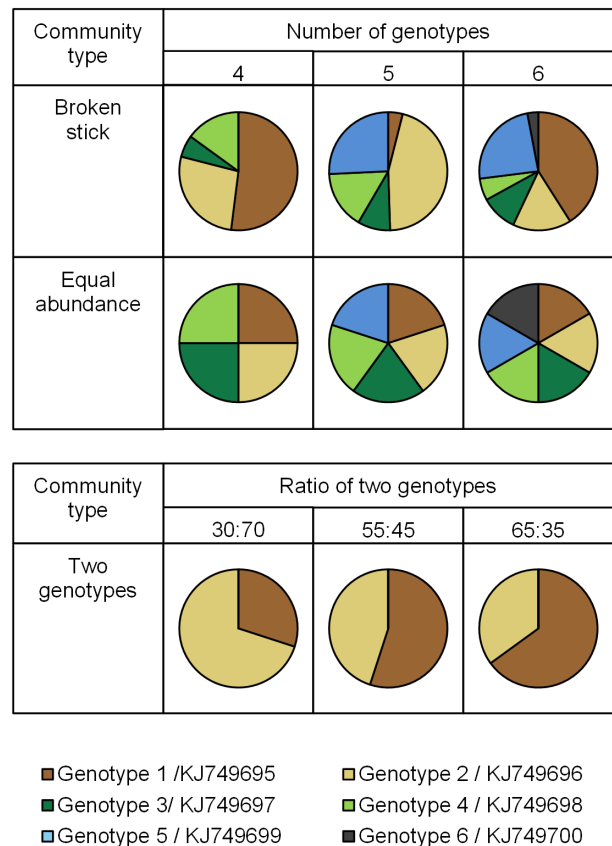
and the concentration of DNA in each was measured in triplicate using a nanodrop-1000 spectrophotometer (NanoDrop Technologies, DE, USA).

**Creating artificial community templates**

Six different artificial community templates were produced by combining four, five or six of the genotypes in different proportions (Figure 1). These different levels of species richness were chosen as these are typical of the number of AM fungal sequence types found in environmental samples (e.g. refs [28,29]). ‘Equal abundance’ templates had all genotypes present in equal amounts, created to test whether particular genotypes were preferentially detected by TRFLP analysis. In ‘broken stick’ templates, the amounts of each component followed a broken stick model, designed to simulate the structure of AM fungal communities often found in environmental samples [26]. Three templates were also produced using just two genotypes in almost equal, but slightly different proportions (Figure 1) to test the resolution of TRFLP analysis. All nine templates were produced in triplicate, with each replicate produced using a different aliquot of its component genotypes.

**Experimental artificial community template analysis using TRFLPs**

All artificial communities were characterised using TRFLPs. This was performed three times on each of the 27 templates to



**Figure 1. Identities and relative proportions of genotypes in each of the nine artificial community templates created.** Accession numbers (KJ749695-KJ749700) refer to the European Molecular Biology Laboratory Nucleotide Sequence Database. doi:10.1371/journal.pone.0109234.g001

provide technical replicates to examine the repeatability of the protocol. To confirm that each genotype produced the predicted TRFs, the three aliquots of each genotype were also analysed. All templates were analysed by amplifying a 550 bp partial fragment of SSU rDNA by PCR using HotstarTaq Plus DNA polymerase (Qiagen) with primers AM1 and NS31 labelled on the 5' end with the fluorescent markers HEX and FAM respectively (Eurofins MWG Operon, Ebersberg, Germany). The 25 µl reactions contained 0.64 mM dNTPs, 0.2 µM of each primer, 0.5 mg dried skimmed milk powder (to prevent PCR inhibition; Premier International Foods Ltd, Lincs., UK.) and the manufacturer's reaction buffer (PCR conditions: 95°C for 5 min; 40 cycles at 95°C for 30 sec, 63°C for 1 min and 72°C for 1 min; and 72°C for 10 min on a Techne TC-512 (Techne Co. Staffs., UK). PCR products were cleaned using a QIAquick PCR Purification Kit (Qiagen) and the concentrations of DNA they contained were measured using a nanodrop-1000 spectrophotometer.

Eight microlitres of each cleaned PCR product were separately digested using five units of *Hinf*-I and *Alu*-I (Promega) in a reaction volume of 15 µl, with manufacturer's buffer and 2 µg BSA for 15 hours. Restriction products were cleaned using a QIAquick PCR Purification Kit (Qiagen). A size standard (GS600-LIZ) was added and samples were run on an ABI 3130 genetic analyzer (Applied Biosystems Inc.). GeneMapper v4.0 software (Applied Biosystems Inc.) was used to determine peak size and quantity. Peaks greater than 100 fluorescent units in height representing TRFs longer than 100 base pairs (bp) were analysed using a bin width of 2 bp and the local southern size calling method. Peak area was used to measure TRF abundance rather than peak height, as using peak height can down-weight longer fragments as they produce wider peaks due to their longer retention time during electrophoresis [30,31]. Before analysis, TRFs which, when present, represented an average of 5% or less of the TRFLP output were excluded to eliminate noise as in ref. [14]. As in ref. [23], only HEX labelled TRFs generated from *Alu*-I digestion and FAM labelled TRFs from *Hinf*-I digestion were further analysed, as other TRFs were predicted to be undetectable and uninformative as they were identical for all genotypes and <50 bp.

## Data analysis

**Predicted TRFLP data.** Theoretical community templates were produced by *in silico* digestion (using *Hinf*-I and *Alu*-I cutting sites) of the six distinct AM fungal genotype sequences. The resulting virtual TRFs were then combined in the same ratios as the artificial community templates. This produced nine predicted theoretical AM fungal communities with which to compare the observed artificial communities empirically analysed using TRFLPs. In addition, the size of possible under-digested terminal fragments, formed where the enzymes did not cleave the DNA at all possible cutting sites (designated 'pseudo-TRFs', see ref. [32]), of each fungal genotype were predicted from their sequences.

**Compositional analysis.** Principal component analysis (PCA) was used to examine whether the TRFLP protocol used could differentiate between the nine artificial communities, when analysed using ordination based on raw species abundance matrices. Covariance PCA was separately performed on the observed data from the empirically tested templates and the predicted data based on *in silico* digestion of the sequences using Minitab v.17 (Minitab Ltd., West Mids., UK). ANOVAs and Tukey's honestly significant difference tests were used to compare the templates' observed PC1 and PC2 scores to determine if they could be differentiated from each other using these community characterisation measures. Identical analyses were also performed on the predicted data.

All possible Euclidean distances between the PC scores of all possible pairs of templates from the observed data were calculated and identical calculations were also performed on the predicted data. To determine the relationship between observed and predicted TRFLP data, the two sets of distances were then compared with a Mantel test (using 10,000 randomisations). Bray Curtis dissimilarity coefficients between pairs of templates were also calculated for all predicted and observed results and similarly compared with a Mantel test, to evaluate TRFLP data as a quantitative measure of community differences.

**Diversity analysis.** Margalef's and Simpson's diversity indices were calculated from both the predicted and observed TRFs to determine how accurately the TRFLP protocol predicted the communities'  $\alpha$  diversity levels. These indices were chosen as Margalef's index is primarily influenced by the richness of a community whereas Simpson's is primarily influenced by evenness. The relationship between predicted and observed indices were tested using linear regression.

Unless stated otherwise, all analyses were conducted using the R statistical language with the 'vegan' and 'ade4' libraries [33–35].

## Results

### Template constituents

There was little variation in length (527 bp–548 bp) of the AM1/NS31 amplicons generated and all sequences had low GC content (39.7%–41.7%), typical of AM fungal AM1/NS31 amplicons. Genotype 2 was similar to sequences from voucher specimens of Gigasporaceae whereas the remaining five genotypes were from the Glomeraceae. The AM1/NS31 amplicon of sequences differed from each other by an average of 9.7% (minimum dissimilarity 4.2%, maximum dissimilarity 16.1%). Genotypes were predicted to produce unique HEX labelled TRFs when digested with enzyme *Alu*-I and unique FAM labelled fragments with enzyme *Hinf*-I (Table 1).

### Detection of TRFs from single genotypes

Observed TRFLP analysis of *Hinf*-I digested amplicons derived from single genotypes showed that an average of 99.2% of the total fragments produced were detected as a single peak, representing the same size fragment as that predicted by the theoretical digests. The sizes of the remaining TRFs corresponded to predicted pseudo-TRFs, which the *in silico* analysis suggested would be produced if the sequences had not fully digested (Table 1). In contrast, instead of producing one peak at approximately the size predicted by the *in silico* analysis, for all but one sequence, digestion of individual genotypes with *Alu*-I produced 2–3 similarly sized peaks of approximately their predicted TRF length, but deviating by between 3.9–13.9 bp (Table 1). When the areas of these peaks were combined, they accounted for 99.9% of the TRFs found in each profile. In all subsequent analysis involving quantifying the relative abundance of these genotypes, the peak areas for these TRFs were therefore combined (Table 1).

### Observed TRFLP data from artificial community templates

The raw TRFLP profiles for observed artificial communities contained 17 TRFs using *Alu*-I (Table 1) whereas the predicted communities contained only six. This discrepancy was mainly caused by the production of multiple peaks for each TRF, and was reduced to eight after pooling based on the digestion of individual genotypes (Table 1). This included the six TRFs predicted from the theoretical analysis, and two TRFs of 310 and 311 bp, which correspond with the size of a predicted pseudo-TRF, formed by

**Table 1.** Predicted and observed TRF sizes using restriction enzymes *Alu-I* and *Hinf-I*.

Enzyme	Genotype identity	Predicted TRF size (bp)	Observed TRF size (bp)	Deviation of observed from predicted TRF size (bp)
<i>Alu-I</i>	Genotype 1 (KJ749695)	162.0	157.7	-4.3
	Genotype 6 (KJ749700)	309.0	302.0/303.8/305.1	-7.0/-5.2/-3.9
			309.6	
			310.6	
	Genotype 4 (KJ749698)	416.0	406.7/411.7	-9.3/-4.3
	Genotype 5 (KJ749699)	449.0	436.0/442.3/443.5	-13.0/-6.7/-5.5
	Genotype 3 (KJ749697)	469.0	455.6/459.5/463.7	-13.4/-9.5/-5.3
	Genotype 2 (KJ749696)	504.0	497.0/498.3/499.2	-7.0/-5.7/-4.8
<i>Hinf-I</i>	Genotype 1 (KJ749695)	141.5	137.6	-3.9
	Genotype 5 (KJ749699)	160.5	154.4	-6.1
	Genotype 4 (KJ749698)	190.5	188.2	-2.3
	Genotype 2 (KJ749696)	299.5	297.4	-2.1
			503.4	
			518.9	
	Genotype 6 (KJ749700)	524.5	523.6	-0.9
			522.6	
Genotype 3 (KJ749697)	547.0	548.1	+1.1	

Genotype identities refer to those given in Figure 1 and accession numbers in the European Molecular Biology Laboratory Nucleotide Sequence Database. Predicted terminal restriction fragment (TRF) sizes are the lengths of fragments (in base pairs; bp) generated by *in silico* digestion of sequences of each genotype. Observed TRFs are those found from the empirical analysis of artificial community templates. TRFs which were produced by empirical TRFLP analysis of individual genotypes and have therefore been combined in further analyses are given in the same row. *Alu-I* TRFs refer to HEX labelled fragments and *Hinf-I* to FAM labelled fragments. doi:10.1371/journal.pone.0109234.t001

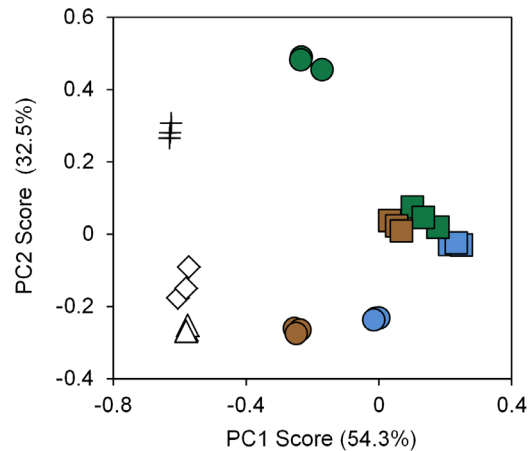
the under-digestion of genotype one, found in all templates (Figure 1; Table 1).

TRFLP analysis of individual genotypes using *Hinf-I* produced nine uniquely sized TRFs. Again, these included the six TRFs predicted from the *in silico* analysis and three fragments (503 bp, 519 bp, 523 bp) which corresponded to the sizes of predicted pseudo-TRFs (Table 1).

**Community differentiation**

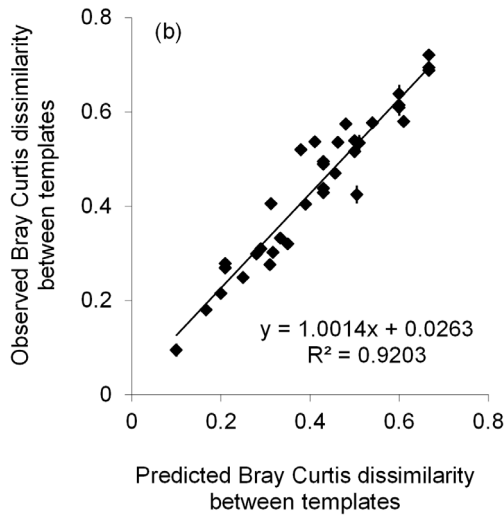
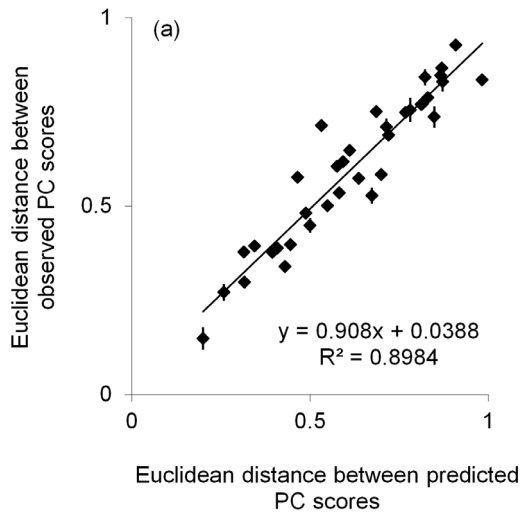
PCA of the predicted and observed TRFLP data explained 88% and 87% of the variation in the data across the first two ordination axes (Figure 2). PC1 and PC2 scores generated from the observed TRF data were significantly different between the nine artificial community templates (Figure 2; ANOVA; PC1 scores:  $F_{8,18} = 761, P < 0.001$ ; PC2 scores:  $F_{8,18} = 465, P < 0.001$ ). Tukey’s honestly significant difference tests showed all templates could be differentiated from each other using either or both their PC1 or PC2 scores ( $P$  values  $< 0.05$ ; Figure 2). Quantitative TRFLP analysis therefore allowed differentiation between all templates with the same number and identity of genotypes, but with different relative abundances. In addition, there was very little variation in PC1 and PC2 scores of the three technical replicates of each of the 27 templates, indicating a high repeatability of this method (Figure 2).

Comparison of predicted and observed Euclidean distances between PC scores across templates produced a significant positive relationship (Figure 3a; Mantel test, using 10,000 randomisations  $r = 0.95, P < 0.001$ ). Comparisons of predicted and observed Bray Curtis dissimilarity coefficients across templates also produced a significant positive relationship (Figure 3b; Mantel test, using 10,000 randomisations;  $r = 0.96, P < 0.001$ ).



- 6 Genotype Broken Stick
- 5 Genotype Broken Stick
- 4 Genotype Broken Stick
- 6 Genotype Equal Abundance
- 5 Genotype Equal Abundance
- 4 Genotype Equal Abundance
- + 2 Genotypes 30:70
- ◇ 2 Genotypes 55:45
- △ 2 Genotypes 65:35

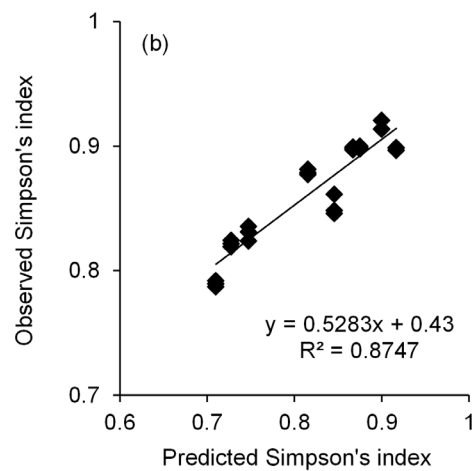
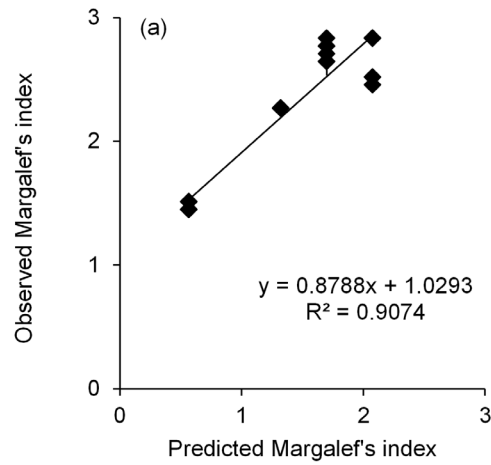
**Figure 2.** Ordination plot of principal component analysis of observed (empirically measured) TRFLP data. Mean principal component factor one (PC1) and two (PC2) scores are plotted for each template replicate. Error bars are standard errors of the means based on technical replicates (replicate PCR and digestion) but they are too small to be visible ( $n = 3$ ). doi:10.1371/journal.pone.0109234.g002



**Figure 3. Relationship between predicted and observed differences between pairs of template communities.** Euclidean distances between all principal component (PC) scores from principal component analysis are shown in (a). Bray Curtis dissimilarity coefficients between pairs of templates are shown in (b). Error bars are standard errors of the means based on analyses of three replicate templates ( $n=3$ ).  
doi:10.1371/journal.pone.0109234.g003

**Diversity measures**

There was a significant linear relationship between predicted and observed Margalef's indices ( $F_{1,25} = 245, P < 0.001, R^2 = 0.91$ ; Figure 4a) showing the TRFLP protocol accurately estimated the differences in richness of the communities from quantitative data. Similarly, there was a significant positive relationship between the predicted and observed Simpson's indices ( $F_{1,25} = 175; P < 0.001, R^2 = 0.87$ ; Figure 4b) indicating quantitative use of the TRFLP data also accurately estimated differences in community evenness.



**Figure 4. Relationship between predicted and observed a) Margalef's indices and b) Simpson's indices.** Error bars are standard errors of the means based on technical replicates (replicate PCR and digestion) but are so small they are rarely visible ( $n=3$ ).  
doi:10.1371/journal.pone.0109234.g004

**Discussion**

**Quantitative use of molecular data for compositional and diversity analyses**

This study demonstrates that the TRFLP protocol tested can be used quantitatively to differentiate among communities of AM fungi with the same taxa but in different proportions (Figure 2). Comparing the predicted and observed TRFLP data using both PCA (Figure 3a) and Bray Curtis dissimilarity coefficients (Figure 3b) produced highly similar results, showing that relative abundance data generated by this technique can be used to accurately quantify relative differences between communities using commonly applied ordination methods. Moreover there was a strong positive relationship between the actual (predicted) and observed community diversities, indicating that both differences in taxon richness (Figure 4a) and evenness (Figure 4b) were accurately recovered. Bias in the different stages of the TRFLP protocol would disrupt these relationships. By empirically testing the limitations of this protocol we thereby demonstrate that quantitative use of TRFLP data is not affected by these biases.

In contrast to some other TRFLP protocols (e.g. ref. [29]), the results of the empirical analysis of experimental artificial community

templates also indicate that the method used produces highly consistent results (as shown by the low variation between technical replicates; Figures 2 & 4). This indicates that stochastic processes affecting the amplification of individual genotypes are unlikely to be important and environmental samples need only be analysed once using this protocol. This highlights that TRFLPs can provide a robust, high throughput, powerful, repeatable method for studying differences in communities of AM fungi.

In addition to showing the strengths and limitations of the TRFLP protocol tested, the findings of this study have implications for other molecular analyses of these fungi. The TRFLP protocol included a PCR step using primers AM1 and NS31, which are also commonly used in other molecular analyses such as pyrosequencing [21,36]. The results of this experiment imply that the amplification of these fungi using these primers was unbiased. This therefore supports the use of data on the relative abundances of AM fungal taxa generated by other methods using these primers. However, it should be noted that confirming this fully is beyond the scope of this study as there are many causes of PCR bias [22] and NGS may have different potential sources [37,38]. The experimental approach described in this paper of generating artificial communities of known characteristics and testing protocols should therefore be adapted and repeated for different methods.

### Studies of individual taxa

Whilst this study suggests that using the abundance of TRFs can be of great use in studying AM fungi at the community level, it also shows extreme caution should be applied when using it to assess the presence and absence of individual taxa. TRFLP analysis overestimated AM fungal richness in all samples (Figure 4a) due to the presence of TRFs that were not expected from the full digestion of the genotypes in the artificial templates (Table 1). These fragments corresponded to the sizes of predicted pseudo-TRFs, defined as non-terminal restriction fragments caused by incomplete digestion [32]. This observed under-digestion was unexpected and difficult to explain. The DNA concentrations of the purified PCR products restricted suggest tenfold more enzyme was used than should have been required for full digestion. Moreover, despite containing the same DNA quantity, the restriction of labeled PCR products of single sequence types exhibited very little under-digestion with over 99% of TRFs produced being the length predicted to be generated by complete digestion. This implies that the formation of pseudo-TRFs occurs more from mixed than single sequence template amplicons. It is possible that pseudo-TRFs were generated by three processes: 1) the presence of chimeric PCR products (i.e. sequences derived from more than one DNA template whereby a combination of two or more sequences are located on the same strand; [22]). This can disrupt restriction, as restriction loci may be swapped between different DNA sequences; 2) single stranded DNA formation as type II restriction enzymes only cleave double stranded DNA [32]; 3) Heteroduplex formation, where DNA molecules form when heterologous sequences bind to each other to form double stranded DNA, which can generate base pair mismatches in the enzyme recognition sequences, reducing efficiency of enzyme digestion [39,40].

The observed pseudo-TRFs show that even when using TRFLPs qualitatively, with 'diagnostic peaks' being used to indicate the presence of a particular genotype in a sample (e.g. as used in ref. [15]), care should be taken to ensure it is impossible to obtain false positives due to incomplete digestion. This could be achieved by *in silico* digestion of a sequence database from the study site, to ensure diagnostic TRFs are not used that are the same size as possible pseudo-TRFs. Moreover, the potential causes of pseudo-TRFs are not restricted to just TRFLP analysis, but happen as a result of errors in PCR. This highlights the

importance of carefully checking richness estimates from any PCR-based method, and may account for the vast range in richnesses reported from NGS studies; different approaches to quality filtering of sequences (e.g. failure to remove all chimeras) may over inflate richness estimates [41]. This study also suggests that ideally, TRFLP analyses should also include empirical testing of the TRF sizes produced from individual fungal genotypes (as conducted in ref. [16]) as restriction of one genotype can unexpectedly produce more than one TRF and the estimated length of fragments can vary slightly from their predicted sizes based on their *in silico* digestions (Table 1), making matching TRFs to sequence databases difficult. This 'TRF drift' is most likely to be due to the effect of different fluorophores on the migratory properties of the fragments during capillary electrophoresis, causing the HEX and FAM labelled TRF fragments to move faster than the size standard, resulting in consistent underestimation of their sizes [42].

### Biological interpretation of molecular AM fungal community characterisation

Finally, when applying the conclusions of this experiment to future studies, standard methodological caveats on the biological interpretation of molecular analyses must also be applied. All TRFLP and NGS analyses measure the relative abundances of fungal taxa within samples, and should therefore not be used to infer differences in absolute abundances of taxa across samples. Furthermore, AM1 and NS31 are thought to amplify the vast majority of AM fungi, but exclude the Paraglomeraceae [43] so cannot be used to determine if there are differences in the communities of this group. Finally, nucleic acid based methods can only be used to assess changes in 'genetic' diversity and composition of AM fungal communities. 'Inter-specific' and temporal variation in gene copy number per unit growth in AM fungi [44,45] mean it is impossible to use them to measure the abundance of fungal 'species' in terms of their biomass. However, as the concept of an AM fungal 'species' is still poorly defined [46], it is arguably preferable to measure these communities in terms of their genetic content anyway.

### Summary

This investigation has shown that TRFLPs can be used quantitatively to discriminate between different AM fungal communities containing the same genotypes in different relative abundances, and accurately measure the differences between them. In addition we have shown that the method is robust and reliable with little error in repeatability. The protocol described therefore has great potential to enhance our understanding of a wide variety of aspects of AM fungal ecology. The results also suggest there was no PCR bias in the protocol, supporting the use of AM fungal relative abundance data from other PCR based methods such as NGS.

### Acknowledgments

We thank Alastair Fitter from the University of York for his helpful comments on this research and Celina Whalley of the University of York Genomics Facility for her technical assistance.

### Author Contributions

Conceived and designed the experiments: TEAC AJD TH. Performed the experiments: TEAC. Analyzed the data: TEAC AJD TH. Contributed reagents/materials/analysis tools: TEAC AJD TH. Wrote the paper: TEAC AJD TH.

## References

- Fitter AH, Moyersoen B (1996) Evolutionary trends in root-microbe symbioses. *Philos Trans R Soc Lond B Biol Sci* 351: 1367–1375.
- Brachmann A, Parniske M (2006) The most widespread symbiosis on earth. *PLoS Biol* 4: e239. DOI: 10.1371/journal.pbio.0040239
- van der Heijden MGA, Klironomos JN, Ursic M, Moutoglou P, Streitwolf-Engel R et al. (1998) Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* 396: 69–72.
- Klironomos JN, McCune J, Hart M, Neville J (2000) The influence of arbuscular mycorrhizae on the relationship between plant diversity and productivity. *Ecol Lett* 3: 137–141.
- Wagg C, Jansa J, Stadler M, Schmid B, van der Heijden MGA (2011) Mycorrhizal fungal identity and diversity relaxes plant-plant competition. *Ecology* 92: 1303–1313.
- Vogelsang KM, Reynolds HL, Bever JD (2006) Mycorrhizal fungal identity and richness determine the diversity and productivity of a tallgrass prairie system. *New Phytol* 172: 554–562.
- Hodge A, Fitter AH (2010) Substantial nitrogen acquisition by arbuscular mycorrhizal fungi from organic material has implications for N cycling. *Proc Natl Acad Sci USA* 107(31): 13754–13759.
- Cheng L, Booke FL, Tu C, Burkey KO, Zhou L et al. (2012) Arbuscular mycorrhizal fungi increase organic carbon decomposition under elevated CO<sub>2</sub>. *Science* 337: 1084–1087.
- Siddiky MRK, Kohler J, Cosme M, Rillig MC (2012) Soil biota effects on soil structure: Interactions between arbuscular mycorrhizal fungal mycelium and collembolan. *Soil Biol Biochem* 50:33–39.
- Drigo B, Pijl AS, Duyts H, Kielak AM, Gamper HA et al. (2010) Shifting carbon flow from roots into associated microbial communities in response to elevated atmospheric CO<sub>2</sub>. *Proc Natl Acad Sci USA* 107: 10938–10942.
- Nuccio EE, Hodge A, Pett-Ridge J, Herman DJ, Weber PK et al. (2013) An arbuscular mycorrhizal fungus significantly modifies the soil bacterial community and nitrogen cycling during litter decomposition. *Environ Microbiol* 15(6): 1870–1881.
- Clapp JP, Helgason T, Daniell TJ, Young JPW (2003) Chapter 8: Genetic studies of the structure and diversity of arbuscular mycorrhizal fungal communities. In: van der Heijden MGA, Sanders IR, editors. *Mycorrhizal Ecology* (Ecological Studies 157). Germany: Springer. pp. 201–221.
- Dickie IA, FitzJohn RG (2007) Using terminal restriction fragment length polymorphism (T-RFLP) to identify mycorrhizal fungi: a methods review. *Mycorrhiza* 17: 259–270.
- Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH (2010) Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J* 4: 337–345.
- Helgason T, Merryweather JW, Young JPW, Fitter AH (2007) Specificity and resilience in the arbuscular mycorrhizal fungi of a natural woodland community. *J Ecol* 95: 623–630.
- Hausmann NT, Hawkes CV (2009) Plant neighborhood control of arbuscular mycorrhizal community composition. *New Phytol* 183: 1188–1200.
- Gosling P, Mead A, Proctor M, Hammond JP, Bending GD (2013) Contrasting arbuscular mycorrhizal communities colonizing different host plants show a similar response to a soil phosphorus concentration gradient. *New Phytol* 198: 546–556.
- Liu WT, Marsh TL, Cheng H, Forney LJ (1997) Characterisation of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* 63(11): 4516–4522.
- Roberts DM, Schofield PG, Don S, Daniell TJ (2012) Directed terminal restriction analysis tool (DRAT): an aid to enzyme selection for directed terminal-restriction fragment length polymorphisms. *Methods Ecol Evol* 3(1): 24–28.
- Öpik M, Metsis M, Daniell TJ, Zobel M, Moora M (2009) Large-scale parallel 454 sequencing reveals host ecological group specificity of arbuscular mycorrhizal fungi in a boreonemoral forest. *New Phytol* 184: 424–437.
- Dumbrell AJ, Ashton PD, Aziz N, Feng G, Nelson M et al. (2011) Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. *New Phytol* 190: 794–804.
- Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng* 96: 317–323.
- Cotton TEA (2010) The effect of atmospheric change on arbuscular mycorrhizal fungal communities. PhD Thesis. University of York, UK.
- Simon L, Lalonde M, Bruns TD (1992) Specific amplification of 18S fungal ribosomal genes from vesicular-arbuscular endomycorrhizal fungi colonizing roots. *Appl Environ Microbiol* 58: 291–295.
- Helgason T, Daniell TJ, Husband R, Fitter AH, Young JPW (1998) Ploughing up the wood-wide web? *Nature* 394: 431.
- Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH (2010) Idiosyncrasy and overdominance in the structure of natural communities of arbuscular mycorrhizal fungi: is there a role for stochastic processes? *J Ecol* 98: 419–428.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876–4882.
- Alguacil MM, Lumini E, Roldán A, Salinas-García JR, Bonfante P, Bianciotto V (2008) The impact of tillage practices on arbuscular mycorrhizal fungal diversity in subtropical crops. *Ecol Appl* 18(2): 527–536.
- Vandenkoomhuyse P, Ridgway KP, Watson IJ, Fitter AH, Young JPW (2003) Co-existing grass species have distinctive arbuscular mycorrhizal communities. *Mol Ecol* 12: 3085–3095.
- Kitts CL (2001) Terminal restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics. *Curr Issues Intest Microbiol* 2: 17–25.
- Sipos R, Székely AJ, Palatinszky M, Révész S, Márialigeti K et al. (2007) Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol Ecol* 60: 341–350.
- Egert M, Friedrich MW (2003) Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. *Appl Environ Microbiol* 69: 2555–2562.
- Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package-I: one-table methods. *R News* 4: 5–10.
- Oksanen J, Blanchet G, Kindt R, Legendre P, Minchin PR et al. (2013) *vegan: Community Ecology Package*. R package version 2.0-7. <http://CRAN.R-project.org/package=vegan>
- R Development Core Team (2007) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>
- Öpik M, Vanatoa A, Vanatoa E, Moora M, Davison J, et al. (2010) The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol* 188: 223–241.
- Berry D, Mahfoudh KB, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl Environ Microbiol* 77(21): 7846–7849.
- Smith DP, Peay KG (2014) Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. *PLoS ONE* 9: e90234. DOI: 10.1371/journal.pone.0090234
- Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res* 30: 2083–2088.
- Petranović M, Petranović D, Dohet C, Brooks P, Radman M (1990) Some restriction endonucleases tolerate single mismatches of the pyrimidine-purine type. *Nucleic Acids Res* 18(8): 2159–2162.
- Fonseca VG, Nichols B, Lallias D, Quince C, Carvalho GR et al. (2012) Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Res* 40(9): e66. DOI: 10.1093/nar/gks002
- Schütte UME, Abdo Z, Bent SJ, Shyu C, Williams CJ et al. (2008) Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Appl Microbiol Biotechnol* 80: 365–380.
- Lee J, Sangsun L, Young JPW (2008) Improved PCR primers for the detection and identification of arbuscular mycorrhizal fungi. *FEMS Microbiol Ecol* 65: 339–349.
- Corradi N, Croll D, Colard A, Kuhn G, Ehinger M et al. (2007) Gene Copy Number Polymorphisms in an Arbuscular Mycorrhizal Fungal Population. *Appl Environ Microbiol* 73(1): 366–369.
- Jansa J, Smith FA, Smith SE (2008) Are there benefits of simultaneous root colonization by different arbuscular mycorrhizal fungi? *New Phytol* 177: 779–789.
- Rosendahl S (2008) Communities, populations and individuals of arbuscular mycorrhizal fungi. *New Phytol* 178: 253–266