Barcoded NS31/AML2 Primers for Sequencing of Arbuscular Mycorrhizal Communities in Environmental Samples

Authors: Morgan, Benjamin S. T., and Egerton-Warburton, Louise M.

Source: Applications in Plant Sciences, 5(8)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1700017

Applications in Plant Sciences

Applications in Plant Sciences 2017 5(8): 1700017

Application Article

Barcoded NS31/AML2 primers for sequencing of arbuscular mycorrhizal communities in environmental samples[1]

Benjamin S. T. Morgan[2,3] and Louise M. Egerton-Warburton[2,3,4]

[2]Program in Plant Biology and Conservation, Northwestern University, Sheridan Road, Evanston, Illinois 60208 USA; and
[3]Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, Illinois 60022 USA

- *Premise of the study:* Arbuscular mycorrhizal fungi (AMF) are globally important root symbioses that enhance plant growth and nutrition and influence ecosystem structure and function. To better characterize levels of AMF diversity relevant to ecosystem function, deeper sequencing depth in environmental samples is needed. In this study, Illumina barcoded primers and a bioinformatics pipeline were developed and applied to study AMF diversity and community structure in environmental samples.
- *Methods:* Libraries of small subunit ribosomal RNA fragment amplicons were amplified from environmental DNA using a single-step PCR reaction with barcoded NS31/AML2 primers. Amplicons were sequenced on an Illumina MiSeq sequencer using version 2, 2 × 250-bp paired-end chemistry, and analyzed using QIIME and RDP Classifier.
- *Results:* Sequencing captured 196 to 6416 operational taxonomic units (OTUs; depending on clustering parameters) representing nine AMF genera. Regardless of clustering parameters, ~20 OTUs dominated AMF communities (78–87% reads) with the remaining reads distributed among other OTUs. Analyses also showed significant biogeographic differences in AMF communities and that community composition could be linked to specific edaphic factors.
- *Discussion:* Barcoded NS31/AML2 primers and Illumina MiSeq sequencing provide a powerful approach to address AMF diversity and variations in fungal assemblages across host plants, ecosystems, and responses to environmental drivers including global change.

Key words: arbuscular mycorrhizal fungi; barcoding; community composition; diversity; Glomeromycota; tropical dry forest.

Arbuscular mycorrhizal fungi (AMF) are a globally important group of fungi that form mutualistic associations with the roots of the majority of land plants (74%; Brundrett, 2009). These mutualisms often improve plant growth and resource acquisition, and thus AMF are recognized as drivers of plant community structure and function and biogeochemical cycling (van der Heijden et al., 1998; Eom et al., 2000). Despite their ecological importance, our understanding of AMF diversity lags far behind that of other groups of fungi (Sanders and Rodriguez, 2016), largely due to the cryptic diversity of phenotypically similar species, late adoption of modern molecular methods, and a study bias toward grassland ecosystems. This limits our understanding of how AMF diversity feeds back to influence plant species' distribution, productivity, diversity, and community assembly (Davison et al., 2012; Bainerd et al., 2014), as well as how AMF–plant interactions vary in response to abiotic drivers such as temperature and precipitation regimes (Egerton-Warburton et al., 2007; Camenzind et al., 2014) and edaphic stresses (e.g., eutrophication, metal-contaminated soils; Cabello, 1997; Bunn et al., 2009; Hassan et al., 2011).

Recent advances in molecular genetic approaches, notably amplicon-targeted high-throughput sequencing technologies such as 454 pyrosequencing, have largely revolutionized the characterization of AMF diversity, phylogeny, and biogeography (Öpik et al., 2009; Schüßler and Walker, 2010) and have revealed high levels of species diversity and complex relationships between AMF and their host plants. For example, these technologies have led to the identification of more than 350 well-characterized molecular taxa (Öpik et al., 2010); shown fine-scale spatial and temporal structuring of AMF communities (Lekberg et al., 2007; Dumbrell et al., 2010, 2011; Bainerd et al., 2014), especially to edaphic constraints (Wang et al., 2016; Wilson et al., 2016); and enabled a sweeping systematic revision of the Glomeromycota (Schüßler and Walker, 2010; Redecker et al., 2013). Although these technologies specifically target AMF species, there is a need to achieve deeper sequencing depth in environmental samples to more completely characterize levels of AMF diversity that are relevant to ecosystem function (Smith and Peay, 2014; Sanders and Rodriguez, 2016).

[1]Manuscript received 27 February 2017; revision accepted 17 June 2017.

The authors thank M. Orvañanos, M. Lazcano Barrero, and L. Leal for site access; M. Öpik for sharing Maarj*AM* resources; and J. Gilbert and S. Owens (Argonne National Laboratory) for technical advice and Illumina sequencing, respectively. This research was funded by a grant to L.M.E.W. from the Institute for Sustainability and Energy at Northwestern University.

[4]Author for correspondence: lwarburton@chicagobotanic.org

doi:10.3732/apps.1700017

Ultra-high-throughput amplicon sequencing capacity using the Illumina MiSeq platform far surpasses 454 technology. It follows that the use of barcoded AMF primers that take advantage of the deep sequencing capacities of the Illumina MiSeq platform could be a promising way to address these needs (Caporaso et al., 2010, 2011). Our objective was to adapt an AMF-specific primer pair (NS31/AML2) with barcodes and to develop and apply a broadly applicable protocol for AMF community amplicon sequencing on the Illumina MiSeq platform. Here, we report on these primers, detail the sequencing protocol and bioinformatics pipeline, and demonstrate the efficacy of our approach by sequencing AMF communities in complex environmental samples from tropical dry forests.

MATERIALS AND METHODS

***Sampling sites and edaphic factors*—**Primers and protocols were tested on environmental samples collected in two seasonally dry tropical forest sites located in the Yucatán Peninsula, Mexico: Reserva Ecológico El Eden (21.195°N, 87.167°W) and Rancho La Higuera (20.445°N, 87.352°W). These sites are privately operated conservation areas (>60 yr after the last disturbance). The local landscape consists of porous Cenozoic limestone (<150 m a.s.l.) overlain by shallow (10–20 cm deep) soils with high organic matter content (20–39% carbon by combustion). The plant community comprises secondary successional forests and is dominated by a high diversity of woody tree species. The climate of the region is warm, subhumid (mean annual temperature 26°C), and strongly seasonal with a wet season (May–October) followed by a marked dry season (November–April). Precipitation ranges from 1100 to 1600 mm/yr, with ~80% of the total precipitation received during the wet season (López-Martinez et al., 2013).

Single soil samples comprising mixed soil and roots (to 20 cm deep) were collected at the drip line of mature *Brosimum alicastrum* Sw., *Vachellia cornigera* (L.) Seigler & Ebinger, *Metopium brownei* (Jacq.) Urb., *Ceiba pentandra* (L.) Gaertn., *Bursera simaruba* (L.) Sarg., and *Manilkara zapota* (L.) P. Royen trees in February 2013 (dry season; $N = 48$). Soil samples were imported into the United States under a U.S. Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) Permit to Receive Soil (P330-11-00358), and all further handling and processing of these samples were conducted in the APHIS authorized containment facility at the Chicago Botanic Garden. Samples were stored frozen (–20°C) until analysis.

Each soil sample was extracted using a ratio of 1 : 10 soil : deionized water, after which the extracts were analyzed for levels of ammonium (NH_4; Weatherburn, 1967), nitrate (NO_3; Doane and Horwath, 2003), phosphate (P; Baykov et al., 1988), and pH and electrical conductivity (EC; Hach Instruments, Loveland, Colorado, USA). In general, soils in La Higuera were of higher pH and contained significantly higher levels of P and NH_4 than those at El Eden (Appendix 1). These differences resulted in a significantly higher soil N : P supply ratio in La Higuera.

***Genomic DNA extraction and amplicon library generation*—**Genomic DNA was extracted from each sample using 0.25 g soil/plant material with the PowerSoil DNA Isolation kit (QIAGEN-MO BIO, Carlsbad, California, USA), following manufacturer's protocols.

Libraries of small subunit ribosomal RNA (hereafter referred to as 18S) fragment amplicons were prepared for each sample with a single-step PCR reaction using primers NS31 (Simon et al., 1992) and AML2 (Lee et al., 2008) that we modified for use with Illumina sequencing platforms following the protocol of Caporaso et al. (2011). These modifications include the addition to both primers of technical adapter sequences for annealing to Illumina flow cells, a standard "pad" sequence, and a novel two-base linker sequence (Table 1); the pad sequence and the two-base linker sequence were both designed to reduce secondary structure formation. Reverse primer constructs were also modified to include a 12-base Golay error-correcting barcode (or index) to enable demultiplexing during data processing (Appendix 2). Primer Prospector (Walters et al., 2011) was used with the Maarj*AM* database of AMF 18S sequences (Öpik et al., 2010) to optimize linker sequences and to test for secondary structure formation in all barcoded primer constructs. Complete sequences of primer constructs NS31f-il and AML2r-il and of all barcodes checked for secondary structure formation are provided in Table 1 and Appendix 2, respectively.

For each sample, PCR was carried out using 10 μL of 5PRIME HotMasterMix, 0.5 μL of NS31f-il, 13 μL of molecular biology–grade water (Fisher Scientific BioReagents, Fair Lawn, New Jersey, USA), 0.5 μL of uniquely barcoded AML2r-il, and 1 μL of genomic DNA extract. The PCR reaction was run using the following thermal cycler program: initial denaturation for 3 min at 94°C; followed by 35 cycles of 45 s at 94°C (denaturation), 60 s at 63.1°C (annealing), and 90 s at 72°C (extension); followed by a final extension step of 10 min at 72°C. PCR reactions were carried out in triplicate for each sample and pooled prior to final sequencing library preparation. The final sequencing library and sequencing reactions were performed at Argonne National Laboratories (Lemont, Illinois, USA). All individual sample amplicon libraries were quantified fluorometrically with a Quant-iT PicoGreen dsDNA assay (Invitrogen Molecular Probes, Eugene, Oregon, USA). An equimolar sequencing library was produced, cleaned using a MO BIO UltraClean PCR Clean-up kit (QIAGEN-MO BIO), and sequenced on an Illumina MiSeq using version 2, 2 × 250-bp paired-end chemistry (Illumina, San Diego, California, USA).

***Data processing and bioinformatics*—**Sequence read processing and bioinformatic analyses (Appendix 3) were performed using QIIME version 1.9.1 (Caporaso et al., 2010), BLAST (Altschul et al., 1990), and the vegan package (Oksanen et al., 2015) in the R statistical environment (R Core Team, 2014). Sequence reads were included in the analyses only if the index read matched a barcode sequence used in this study with two or fewer errors.

During data processing, we found that a large proportion of sequenced amplicons were too long to allow for overlap with the Illumina MiSeq version 2, 2 × 250-bp sequencing technology, and thus could not be aligned and assembled (Appendix 4). The average amplicon length was ~530 bp based on alignment of paired forward and reverse reads to full-length amplicon regions in the Maarj*AM* Virtual Taxon (VT) database of 348 well-resolved AMF species-level sequence clusters (http://maarjam.botany.ut.ee). Instead, we conducted all downstream

TABLE 1. Individual components and complete sequences of Illumina MiSeq-compatible custom sequencing primers used to sequence amplicon libraries in this study and their estimated melting temperatures.

PCR direction and primers	Primer sequences (5′–3′)	T_m (°C)
Forward		
5′ Illumina adapter P5	AATGATACGGCGACCACCGAGATCTACAC	—
Forward primer pad	TATGGTAATT	—
Forward primer linker	CT	—
Forward primer (NS31)	TTGGAGGGCAAGTCTGGTGCC	—
Complete forward primer construct (NS31f_il)	AATGATACGGCGACCACCGAGATCTACACTATGGTAATTCTTTGGAGGGCAAGTCTGGTGCC	70.6
Reverse		
Reverse complement of 3′ Illumina adapter P7	CAAGCAGAAGACGGCATACGAGAT	—
Golay barcode (see Appendix 2)	XXXXXXXXXXXX	—
Reverse primer pad	AGTCAGTCAG	—
Reverse primer linker	AC	—
Reverse primer (AML2)	GAACCCAAACACTTTGGTTTCC	—
Complete reverse primer construct (AML2r-il)	CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXXXXAGTCAGTCAGACGAACCCAAACACTTTGGTTTCC	69.8–71.3

Note: T_m = melting temperature.

TABLE 2. Total number of operational taxonomic units (OTUs) clustered at three similarity thresholds showing the number of OTUs retained after each filtering step during data processing, and used in each of the nine analyzed data sets.

	Clustering threshold		
Factor	90%	95%	97%
Total OTUs	5975	46,066	102,255
Passed BLAST vs Maarj*AM*	399	2857	7288
Passed UCHIME	395	2836	7279
Passed BLAST vs NCBI nt	385	2795	7255
No. OTUs in all OTUs data set	365	2524	6416
No. OTUs in ≥2-ton data set	276	1250	2213
No. OTUs in ≥10-ton data set	196	374	407

Note: NCBI nt = National Center for Biotechnology Information nonredundant nucleotide database; OTU = operational taxonomic unit.

analyses using only the 250-bp forward reads based on evidence provided by Davison et al. (2012). More specifically, Davison et al. (2012) found that artificially truncating NS31/AML2 reads from 400 to 170 bp resulted in a near identical capacity to capture AMF diversity because the majority of taxonomically informative characters occurred in the 5′-most 170 bases.

Raw reads were demultiplexed and quality filtered with default parameters and a quality threshold of 20. Sequencing reads were truncated after four consecutive base calls with quality scores less than 20 (99% confidence interval). Truncated reads that were less than 75% of their original length were removed from further analyses. Because input reads were 250 bp, the resulting data set contained reads with a minimum 187 bp.

Operational taxonomic units (OTUs) were clustered using reads from all 48 AMF community libraries using an open reference strategy and the Maarj*AM* VT database. We clustered OTUs at 97% similarity, which is conventionally used as a species-level threshold, and also at 95% and 90% similarities, as numerous reports have documented intraspecific, and even intraindividual, variation in AMF ribosomal DNA that exceeds the 3% dissimilarity threshold (e.g., Clapp et al., 1999; Rodriguez et al., 2004).

Probable artifacts and sequences from nontarget taxa were excluded by BLAST against the complete Maarj*AM* database of 5934 AMF 18S sequences. OTUs were excluded if they did not hit a Maarj*AM* database sequence with an *E*-value below 10^{-50} and alignment of at least 90% of the read length with sequence identity equal to or greater than the OTU clustering similarity threshold (e.g., 90% sequence identity for 90% OTUs). We also used a second BLAST search against a database of primer constructs to screen for OTUs that may have incorporated any sequence originating from the primer (*E*-value threshold of 10^{-10}; identity threshold of 50%), but no further OTUs were removed in this step. Chimeric OTUs were identified using the USearch v6 implementation of UCHIME (Edgar et al., 2011) with default parameters and a database of all 5934 sequences from the Maarj*AM* database and their reverse complements. Any OTU with a UCHIME score greater than 1 was removed from further analyses. Finally, all OTU-type sequences were BLASTed against the National Center for Biotechnology Information (NCBI) nonredundant nucleotide database, and any OTU with a best hit that was not identified as an AMF in the GenBank record was removed from further analyses. Taxonomy was assigned to remaining OTUs using the QIIME implementation of RDP Classifier (Wang et al., 2007) retrained using the complete Maarj*AM* database, and with a minimum confidence threshold of 0.8. Finally, to reduce sampling imbalance between sites, samples with fewer than 1000 reads remaining after all quality-filtering steps were removed. This approach resulted in a total of 33 samples, representing 14 samples from El Eden and 19 samples from La Higuera.

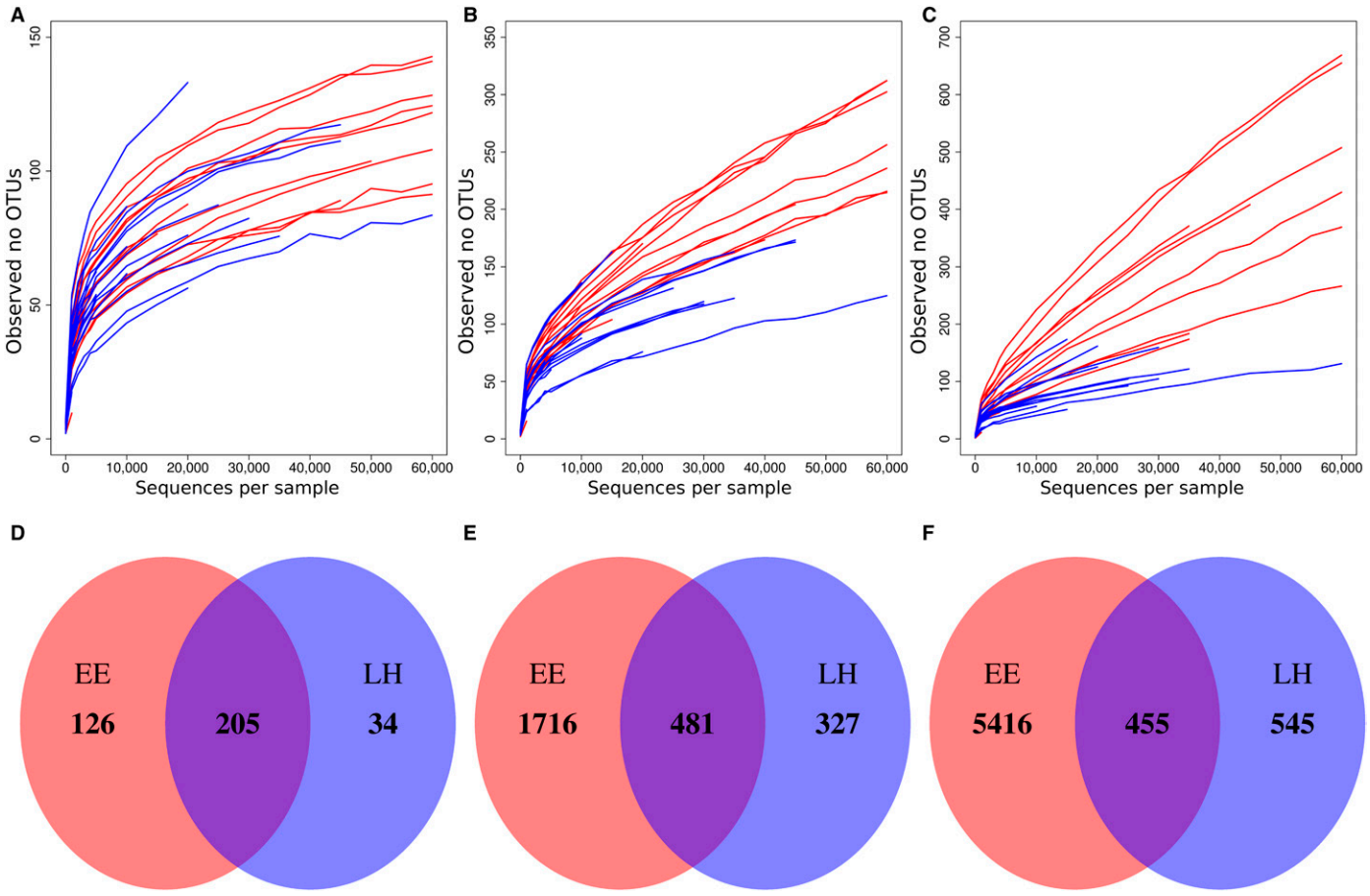


Fig. 1. Alpha rarefaction curves for 90% similar (A), 95% similar (B), and 97% similar (C) operational taxonomic units (OTUs) in samples from El Eden (red) and La Higuera (blue); Venn diagrams illustrating the number of 90% similar (D), 95% similar (E), and 97% similar (F) OTUs unique to El Eden (EE) or La Higuera (LH), or present at both sites. Data are shown for the most inclusive, all OTUs data sets.

For each collection of OTUs (El Eden, La Higuera), we explored the effects of rare taxa on patterns of diversity by conducting three separate sets of analyses at both sites that included all OTUs, excluded singletons, or excluded all OTUs with fewer than 10 constituent sequences to capture only core diversity (Smith and Peay, 2014). The removal of rare OTUs did not result in read numbers falling below the 1000-reads/sample threshold. Table 2 demonstrates the number of OTUs retained in each clustering threshold following the quality-filtering steps.

We used QIIME to generate Bray–Curtis dissimilarity matrices for each data set using rarefied data (1000 reads/sample) and tested for significant differences in AMF community structure between sites and in relation to soil chemical variables (log-transformed except pH) using principal coordinates analysis (PCoA) and PERMANOVA tests with the capscale and ADONIS functions in the vegan R package. Next, we tested whether there were significant differences in AMF community dispersion among sites using the betadisper function in vegan. We used QIIME to generate $Chao_1$, Shannon–Wiener, and Simpson (1-D) alpha diversity metrics for each sample (not rarefied). Differences in OTU richness and diversity and soil chemical variables between sites were analyzed using ANOVA and Tukey's honest significant difference posthoc tests in R. Finally, we used the vegdist function in vegan to generate a Euclidean distance matrix based on soil chemical variables (NO_3, NH_4, P, N : P supply ratio, EC, pH), and used a Mantel test with Pearson and Spearman coefficients to test for correlation between the soil chemical and the AMF community matrices.

RESULTS

The 5,977,389 quality-filtered, demultiplexed sequence reads used in this study have been uploaded to the NCBI Sequence Read Archive, and are associated with BioProject PRJNA329250.

***General primer performance*—**After quality filtering and exclusion of reads belonging to artifacts, nontarget taxa, and rare OTUs (if any), each of the nine data sets contained a total number of reads ranging from 2,325,440 (97%, ≥10-ton set) to 2,776,849 (90%, all OTUs set). Mean reads per sample ranged from 70,468 ± 116,616 (mean ± SD) to 84,146 ± 133,491, while the median number of reads for samples in a data set ranged from 22,932 to 28,368. Alpha rarefaction indicated that ~20,000 to >60,000 reads per sample may be necessary to adequately sample these AMF communities, with more reads required at higher similarity thresholds (Fig. 1A–C). This was particularly notable in the El Eden data sets (Fig. 2) and in more rare-OTU-inclusive data (Appendix S1).

***AMF species identification*—**The total number of OTUs increased markedly with increasing clustering similarity threshold and rare OTU inclusivity in both sites (Fig. 1D–F), and ranged from 196 (90% similar ≥10-ton OTUs) to 6416 (97% similar all OTUs; Table 2).

RDP Classifier assigned 91.5% (90% similarity), 99.2% (95% similarity), and 99.8% (97% similarity) of OTUs to genus-level or species-level accessions in the VT database with at least 80% confidence. Overall, these OTUs accounted for >99.5% of the sequencing reads in each data set.

AMF communities were characterized by a small group of highly abundant AMF species and numerous rare species. In each clustering threshold, AMF communities were dominated by 23–25 OTUs that, in total, represented ~78–87% of reads. The remaining reads were distributed among the numerous subordinate (rare) OTUs (Appendix S2). This pattern is consistent with the lognormal model distribution of AMF species noted in previous studies (Dumbrell et al., 2010) and suggests that a large number of specialist and/or narrowly endemic AMF fungi occurred in these forests.

Nine genus-level assignments were made to OTUs and were consistent with well-characterized AMF genera from community studies (Redecker et al., 2013). As in other studies of AMF communities (Eom et al., 2000; Egerton-Warburton et al., 2007), *Glomus* sensu lato (s.l.) (taxa formerly classed as *Glomus* Group A) dominated the AMF community and accounted for 76–97% of OTUs and >90% of all sequence reads in every data set (Table 3; Fig. 3; Appendix S3).

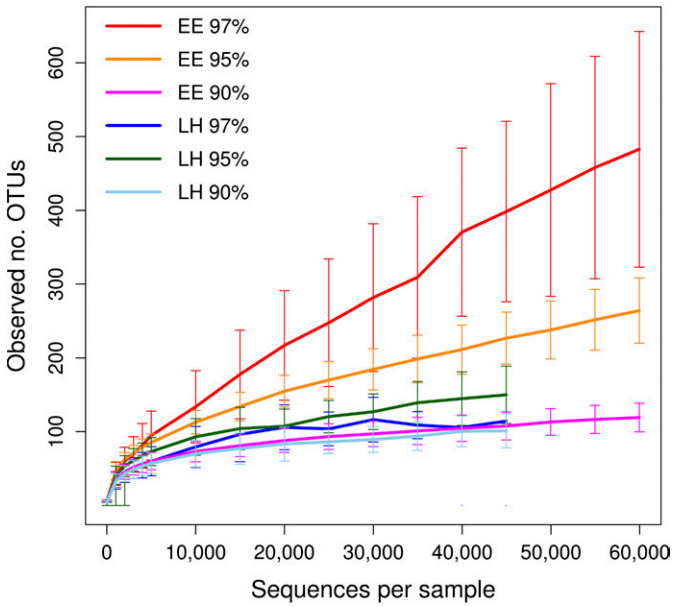


Fig. 2. Averaged alpha rarefaction curves of observed operational taxonomic unit (OTU) richness at 90%, 95%, and 97% similarity clustering thresholds in El Eden (EE) and La Higuera (LH). Vertical bars represent the standard deviation of the mean.

RDP Classifier assigned 2–8% of OTUs in each clustering threshold to a species-level VT. The majority of species-level assignments were to *Glomus* VT; this genus comprised 54%, 70%, and 94% of AMF communities in 90%, 95%, and 97% clustering thresholds, respectively. In each clustering threshold, at least one species-level assignment was made in each genus except *Gigaspora* (Table 3). Complete RDP Classifier assignments for all OTUs at each clustering similarity threshold are provided in Appendices S4, S5, and S6.

***AMF community diversity, structure, and composition*—**Our analyses revealed strong positive effects of both increasing OTU clustering similarity threshold and increasing rare OTU

TABLE 3. Number of operational taxonomic units assigned to each of nine arbuscular mycorrhizal fungal genera using three clustering similarity thresholds.

Genus	Clustering threshold		
	90%	95%	97%
Glomus Tul. & C. Tul. s.l.	277	2392	6229
Diversispora C. Walker & A. Schüßler	20	71	66
Claroideoglomus C. Walker & A. Schüßler	13	1	2
Paraglomus J. B. Morton & D. Redecker	10	19	15
Acaulospora Gerd. & Trappe	6	2	4
Scutellospora C. Walker & F. E. Sanders	4	14	79
Ambispora C. Walker, Vestberg & A. Schüßler	1	2	1
Archaeospora J. B. Morton & D. Redecker	2	3	2
Gigaspora Gerd. & Trappe	1	1	2
Genus could not be assigned	31	19	16

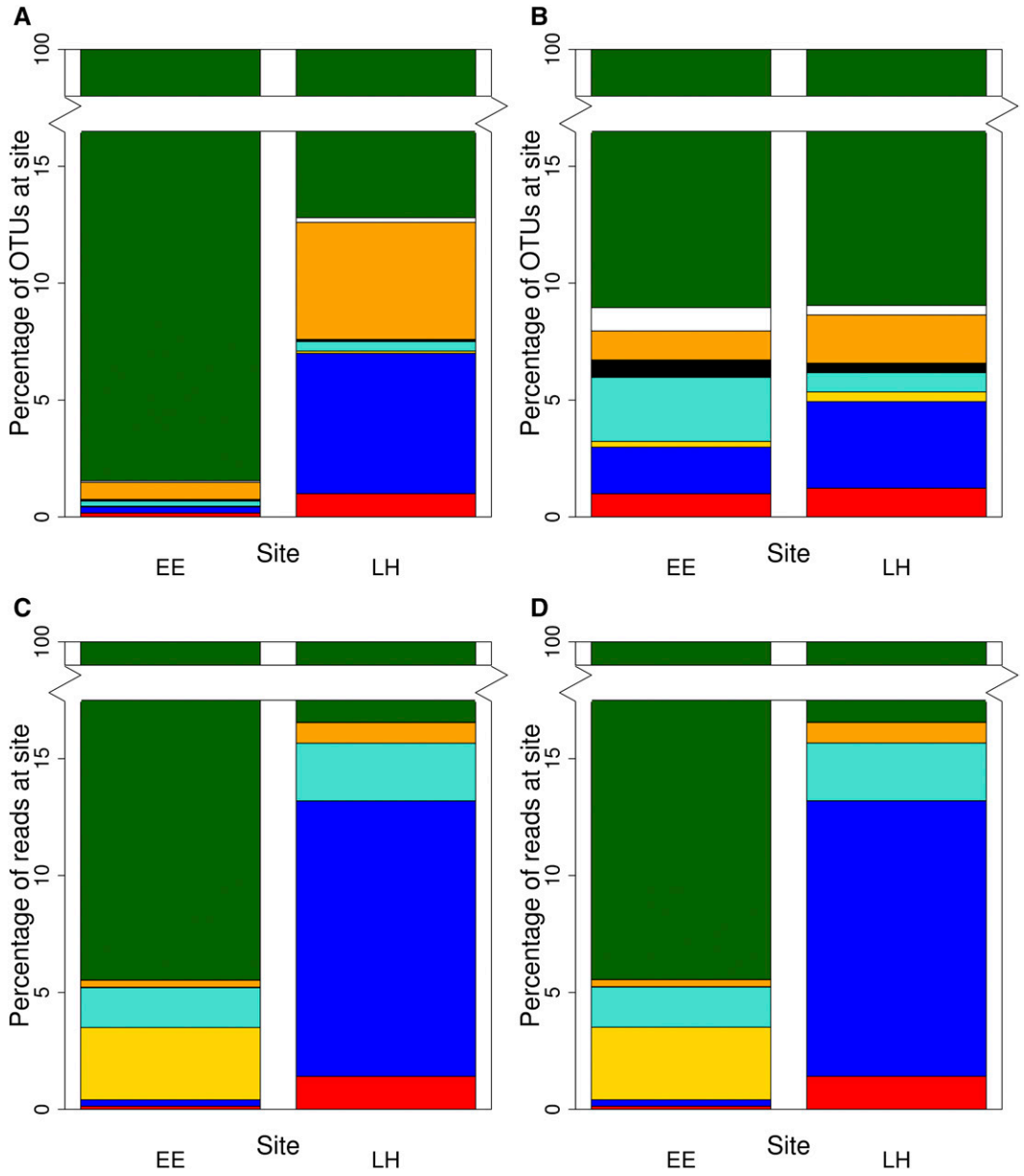


Fig. 3. The percentage of operational taxonomic units (OTUs) (A, B) and percentage of sequence reads (C, D) assigned to arbuscular mycorrhizal fungal genera at El Eden (EE) and La Higuera (LH) study sites. Data are shown for 97% similar OTUs at two inclusivity levels: all clustered OTUs (A, C) and ≥10-ton OTUs (B, D). Colors indicate genus assigned by RDP Classifier: *Glomus* (green); *Ambispora*, *Archaeospora*, and *Gigaspora* (white); *Scutellospora* (orange); *Acaulospora* (black); *Paraglomus* (aqua); *Claroideoglomus* (yellow); *Diversispora* (blue); and no genus-level assignment (red).

inclusivity on observed and estimated taxonomic richness (Tables 2, 4). Specifically, increasing the OTU clustering similarity threshold significantly inflated the observed OTU richness owing to the clustering of rare OTUs (singletons, <10 tons) in the most read-rich samples in El Eden (Table 4; Appendices S7, S8, and S9). Even so, there were consistent patterns of OTU richness and diversity across all data sets. For example, OTU richness (observed, $Chao_1$ estimates) was significantly higher in El Eden than La Higuera ($P < 0.02$) while Shannon–Wiener and Simpson index values did not differ significantly between sites (Table 4). In addition, OTU richness was negatively correlated with soil pH ($P < 0.020$) and NH_4 levels ($P < 0.028$) in all data sets.

AMF community composition was also impacted by OTU clustering threshold and rare OTU inclusivity. However, some consistent trends were apparent. For example, AMF communities were dominated by species of *Glomus* s.l., and those in La Higuera contained a greater abundance and diversity of *Diversispora* species and lower levels of *Claroideoglomus* than El Eden (Fig. 3). Removing the rare OTUs from analyses increased the similarity in AMF community composition between sites with respect to the proportion of OTUs assigned to each genus (Fig. 3A, B; Appendix S3) but had little effect on read abundances (Fig. 3C, D). These results support the presence of site-specific AMF taxa and suggest that the primary difference in proportion of OTUs assigned to each genus among sites was due to relatively

TABLE 4. Observed operational taxonomic unit (OTU) richness per individual sample and site, number of OTUs unique to each site, and indices of diversity ($Chao_1$, Simpson, Shannon–Wiener) for arbuscular mycorrhizal fungal communities at both study sites. Values represent the site mean with standard deviation in parentheses; El Eden, $N = 14$ samples; La Higuera, $N = 19$ samples.

			No. of OTUs*			Diversity indices*		
Clustering threshold	OTUs	Site	Per sample	Total	Unique	$Chao_1$	Simpson	Shannon–Wiener
90%	All	El Eden	120 (12)[a]	331	126	160 (15)[a]	0.79 (0.050)[a]	3.24 (0.25)[a]
		La Higuera	79 (15)[b]	239	34	100 (20)[b]	0.68 (0.066)[a]	2.74 (0.33)[a]
	≥2 ton	El Eden	116 (11)[a]	267	62	144 (13)[a]	0.79 (0.050)[a]	3.24 (0.25)[a]
		La Higuera	78 (15)[b]	214	9	96 (17)[b]	0.68 (0.066)[a]	2.74 (0.33)[a]
	≥10 ton	El Eden	106 (10)[a]	194	19	126 (10)[a]	0.79 (0.050)[a]	3.24 (0.25)[a]
		La Higuera	75 (13)[b]	177	2	91 (14)[b]	0.68 (0.066)[a]	2.74 (0.33)[a]
95%	All	El Eden	337 (44)[a]	2197	1716	718 (94)[a]	0.84 (0.032)[a]	3.76 (0.22)[a]
		La Higuera	122 (49)[b]	808	327	204 (123)[b]	0.83 (0.042)[a]	3.44 (0.29)[a]
	≥2 ton	El Eden	264 (31)[a]	1178	697	409 (42)[a]	0.84 (0.032)[a]	3.75 (0.22)[a]
		La Higuera	109 (41)[b]	553	72	154 (55)[b]	0.83 (0.042)[a]	3.44 (0.29)[a]
	≥10 ton	El Eden	121 (14)[a]	371	89	186 (15)[a]	0.84 (0.032)[a]	3.73 (0.22)[a]
		La Higuera	90 (18)[b]	285	3	113 (20)[b]	0.83 (0.042)[a]	3.43 (0.29)[b]
97%	All	El Eden	617 (103)[a]	5871	5416	1845 (329)[a]	0.79 (0.046)[a]	3.48 (0.26)[a]
		La Higuera	108 (135)[b]	1000	545	258 (434)[b]	0.82 (0.060)[a]	3.23 (0.34)[a]
	≥2 ton	El Eden	349 (50)[a]	2219	1674	533 (71)[a]	0.79 (0.046)[a]	3.45 (0.26)[a]
		La Higuera	84 (66)[b]	539	84	132 (94)[b]	0.82 (0.060)[a]	3.22 (0.34)[a]
	≥10 ton	El Eden	130 (13)[a]	402	164	160 (15)[a]	0.79 (0.046)[a]	3.41 (0.26)[a]
		La Higuera	65 (18)[b]	243	5	88 (20)[b]	0.82 (0.060)[a]	3.21 (0.34)[a]

*For each clustering threshold and OTU inclusivity level, means within each column with the same letter do not differ significantly at $P < 0.05$ (Tukey's HSD test).

large numbers of rare taxa (i.e., rare *Diversispora* OTUs at La Higuera), while differences in read abundance assigned to genera between sites were due to a few highly abundant OTUs.

These results were further supported by significant differences in AMF community structure ($P < 0.0047$, PERMANOVA; Fig. 4), but no significant difference in dispersion between sites ($P > 0.28$ for all data sets, PERMDISP2). AMF community structure was significantly influenced by pH ($P < 0.0035$, PERMANOVA) and NH_4 ($P < 0.0062$, PERMANOVA; Fig. 4) because OTU richness was negatively influenced by increasing soil pH or NH_4 level (Fig. 5). These patterns were consistent across clustering thresholds (Appendix S10). Comparisons between PCoA ordinations for environmental variables (Euclidean distance; Fig. 6A) and OTUs at all clustering thresholds and inclusivity levels (Bray–Curtis distance; 90% similar, ≥2-ton data shown in Fig. 6B) support a highly significant correlation between AMF community composition and soil properties using either Pearson or Spearman coefficients (Mantel test: $P < 0.004$, $R^2 > 0.20$).

Data sets using the three OTU clustering similarity thresholds also showed that AMF community structure was responsive to soil P levels (Appendix S10). Negative relationships between taxonomic richness and soil P levels were only supported

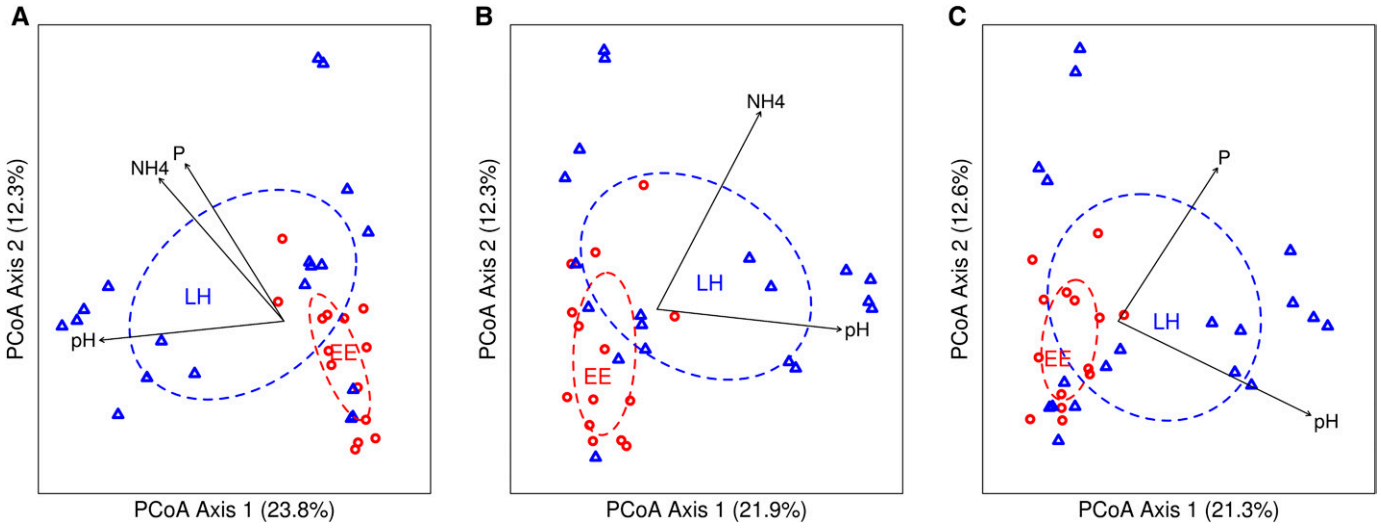


Fig. 4. Principal coordinates analysis (PCoA) ordination plots of arbuscular mycorrhizal fungal (AMF) communities sampled at El Eden (EE, red circles) and La Higuera (LH, blue triangles) using Bray–Curtis dissimilarities based on all operational taxonomic units (OTUs) clustered at 90% (A), 95% (B), and 97% (C) similarity thresholds. Percentage values on the axes represent the variation in AMF community dissimilarity explained by each axis. Ellipses represent the central tendency of communities at each site. Vectors denote the magnitude and direction of statistically significant effects of soil properties on AMF community dissimilarity.

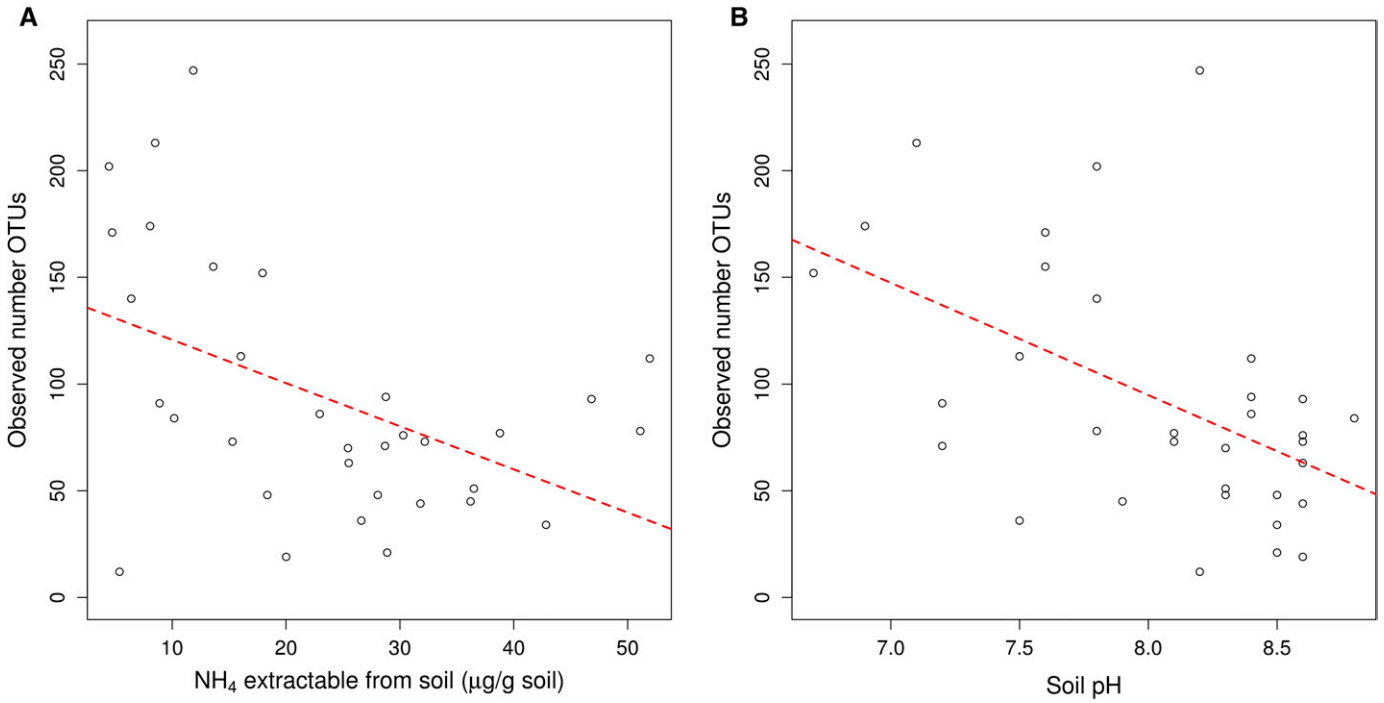


Fig. 5. Relationship between operational taxonomic unit (OTU) richness and soil NH_4 (A) and pH (B) in the 97% ≥10-ton data set, which is representative of patterns observed across all data sets. Red dashed lines show the best fit of linear regression models with $P < 0.003$.

($P < 0.05$, ANOVA) in 95% and 97% similar data sets, while AMF community structure responses to soil P were only supported in 90% and 97% similar data sets ($P < 0.030$, PERMANOVA). Rare OTU inclusivity had little effect on these patterns.

DISCUSSION

The overarching goal of our study was to determine whether AMF diversity and variations in AMF communities could be adequately captured on the Illumina MiSeq platform, and to determine its potential utility in large-scale surveys of AMF communities. To address this goal, we modified existing 18S primers for barcoding, applied robust protocols with which to undertake 18S amplicon analysis on the Illumina MiSeq platform, and developed bioinformatics pipelines for data processing.

Our results clearly demonstrate that the application of barcoded NS31/AML2 primers improves the level of resolution in AMF species identification, diversity, and community composition in complex environmental samples. This primer pair effectively amplified a wide diversity of AMF genera and species, and did not appear to exclude taxa that have been omitted previously due to primer bias (e.g., Archaeosporaceae and Paraglomeraceae; Lee et al., 2008). Along with the deeper sequence coverage provide by the Illumina MiSeq, this approach also revealed one of the highest levels of AMF species richness recorded to date (Öpik et al., 2010), ranging from 196 OTUs in the most conservative data set (≥10-ton, 90% threshold) to more than 6000 at the highest levels of OTU clustering similarity and rare OTU inclusivity (all samples, 97% threshold). A large percentage of AMF taxa were also present in extremely low abundance, thereby supporting the capacity of our protocols to capture rare taxa. In contrast, previous studies using 454 pyrosequencing indicated that AMF communities hosted, on average, 70 AMF taxa (e.g., Dumbrell et al., 2010), while estimates using morphological methods indicated ~45 AMF taxa within a community (Eom et al., 2000; Egerton-Warburton et al., 2007).

We also captured biogeographic differences in AMF communities between the two study sites. Across all data sets, there were consistent and significant differences in OTU richness abundance. For example, La Higuera contained a greater abundance and diversity of *Diversispora* species and lower levels of *Claroideoglomus* than El Eden. In addition, the significantly higher pH and levels of NH_4 and P at La Higuera appeared to drive the observed site effect on AMF community composition and structure. These results are in general agreement with spore-based studies of AMF communities in other systems (Egerton-Warburton et al., 2007). Our study was not designed to examine the mechanistic basis of these results. Based on studies elsewhere, however, it is possible that the dry season constrains the AMF community to taxa that are physically or physiologically tolerant of low soil moisture (e.g., *Glomus* and *Diversispora*; Augé, 2001) or to high soil P fertility or pH (Wang et al., 2016). Alternatively, these shifts may reflect differences in host plant requirements during the dry season for AMF that increase host drought tolerance (e.g., stomatal control, cytokinin production; Augé, 2001) or increase the acquisition of limiting nutrients from carbonate substrates (N, Fe, Zn; Labidi et al., 2012). Irrespective, our results indicate a high potential to use our AMF protocol in large-scale sequencing projects to address AMF diversity with sufficient taxonomic precision, to determine the extent to which AMF assemblages vary across host plants and ecosystems, and to resolve AMF species' responses to edaphic stressors, such as anthropogenic N deposition, in complex environmental samples.

Our study also highlighted a number of technical considerations. First, a key finding was OTU inflation, particularly at the 97% clustering threshold, which is the level applied in most

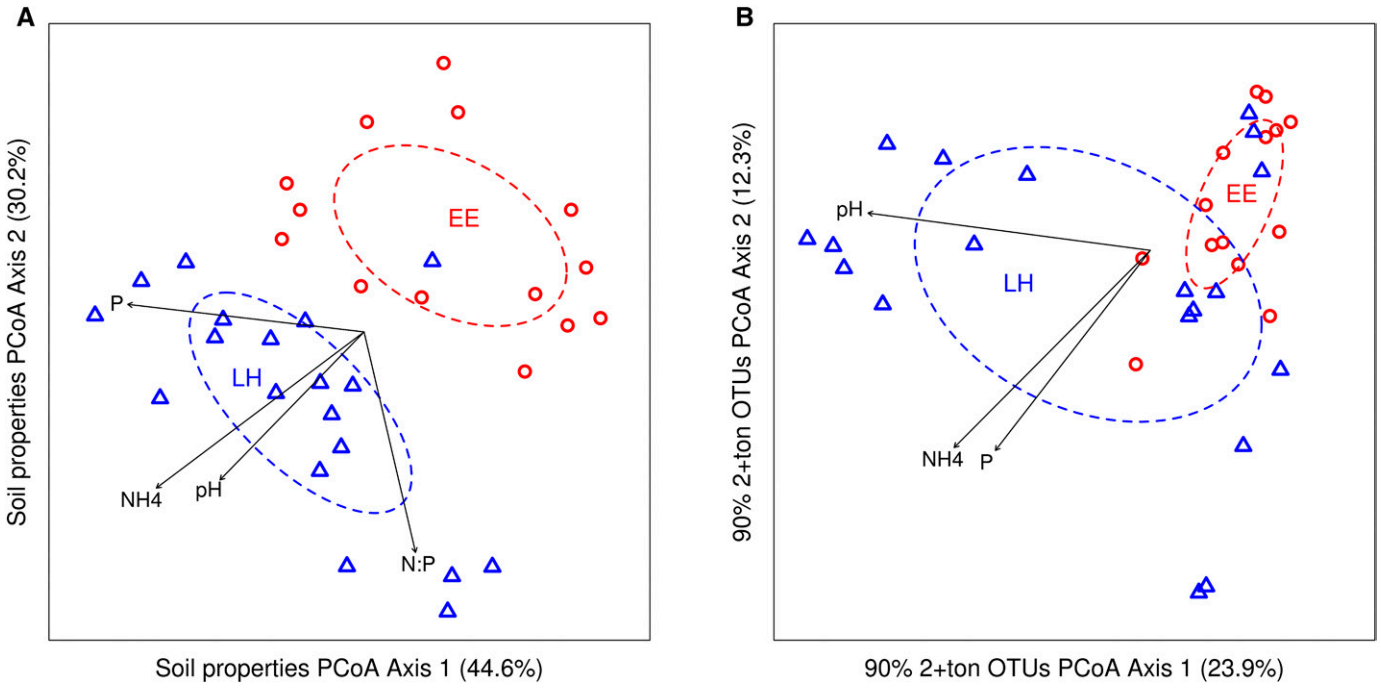


Fig. 6. Principal coordinates analysis (PCoA) ordination plot of soil properties (A) and arbuscular mycorrhizal fungal (AMF) communities (B) in El Eden (EE) and La Higuera (LH). Percentage values on the axes represent the variation in soil properties (A) and AMF operational taxonomic unit (OTU) read abundance (B) explained by each axis. Ellipses represent the central tendency of communities, and vectors denote the magnitude and direction of the effects of significant soil nutrients on AMF communities.

mycorrhizal fungal studies. Between the 97% and 90% clustering thresholds, there was a twofold increase in OTU richness at the ≥10-ton level, an 8-fold increase when considering ≥2-ton OTUs, and a 17-fold increase when considering all clustered OTUs (Table 2). In our study, this result was primarily due to the recovery of numerous rare or unique AMF taxa in the most read-rich samples (see Table 4; Appendix S1), rather than to issues such as uneven number of sequences among samples. To avoid overestimation of AMF community diversity, excluding all OTUs with fewer than 10 constituent sequences (regardless of clustering threshold) will result in levels of taxonomic (OTU) richness consistent with current estimates of AMF species richness (Öpik et al., 2010; Schüßler and Walker, 2010).

Second, the majority of OTUs could not be assigned to any species-level accession in the Maarj*AM* database. While this result indicates that novel AMF species likely occur in the Yucatán as they do in other tropical systems (Chaiyasen et al., 2014), it raises questions about our ability to identify AMF species and catalog their diversity in environmental samples. One possibility is that OTU matching was hampered by the limited availability of well-characterized AMF taxa from tropical forests. Alternatively, the clustering of sequences to generate AMF VT (Öpik et al., 2010) may have reduced the potential for OTU matching if sequences of multiple species were lumped into the same OTU (Bruns and Taylor, 2016) or if sequences from a single species were assigned to multiple OTUs (House et al., 2016). Either scenario could mask phylogenetic diversity (and inferences of functionality) or AMF species with large amounts of intraspecific variation. Thus, a more comprehensive reference database of AMF sequences is now needed to improve our capacity to identify AMF to species level and to address within-species sequence variation (House et al., 2016). In particular, more direct assessments that use well-characterized sequences from individual spores or single-spore cultures are needed.

Finally, there are limitations to using MiSeq version 2, 2 × 250-bp chemistry with AMF-barcoded samples. Improvements in the MiSeq chemistry with version 3 (2 × 300 bp) is expected to further improve the differentiation of OTUs and the taxonomic resolution of AMF species by allowing consistent assembly of forward and reverse reads into a ~530-bp sequence fragment. Preliminary analyses of recent 2 × 300-bp sequencing data suggest that OTUs clustered using assembled reads are assigned to species level at approximately two times the rate of OTUs clustered from forward reads alone using similar screening stringency levels (Morgan and Egerton-Warburton, unpublished data).

Our results support the continuing development and use of high-throughput sequencing approaches to address the AMF "black box." As a first step, the tools detailed herein allow the detection of ecologically relevant levels of AMF diversity that shape plant community composition, diversity, and nutrient acquisition in natural and restored communities, including rare and unique species (Sanders and Rodriguez, 2016). While this approach is broadly applicable to most ecosystems, it is especially important in those where major gaps remain in our understanding of AMF species richness and their role in plant community composition and function.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Augé, R. M. 2001. Water relations, drought and vesicular-arbuscular mycorrhizal symbiosis. *Mycorrhiza* 11: 3–42.

Bainerd, L., J. Bainerd, C. Hamel, and Y. Gan. 2014. Spatial and temporal structuring of arbuscular mycorrhizal communities is differentially influenced by abiotic factors and host crop in a semi-arid prairie agroecosystem. *ISME Journal* 88: 333–344.

Baykov, A., O. A. Evtushenko, and S. M. Avaeva. 1988. A malachite green procedure for orthophosphate determination in alkaline-phosphatase-based enzyme immunoassay. *Analytical Biochemistry* 171: 266–270.

Brundrett, M. C. 2009. Mycorrhizal associations and other means of nutrition of vascular plants: Understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis. *Plant and Soil* 320: 37–77.

Bruns, T. D., and J. W. Taylor. 2016. Comment on 'Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism'. *Science* 351: 826.

Bunn, R., Y. Lekberg, and C. Zabinski. 2009. Arbuscular mycorrhizal fungi ameliorate temperature stress in thermophilic plants. *Ecology* 90: 1378–1388.

Cabello, M. N. 1997. Hydrocarbon pollution: Its effect on native arbuscular mycorrhizal fungi (AMF). *FEMS Microbiology Ecology* 22: 233–236.

Camenzind, T., S. Hempel, J. Homeier, S. Horn, A. Velescu, W. Wilcke, and M. C. Rillig. 2014. Nitrogen and phosphorus additions impact arbuscular mycorrhizal abundance and molecular diversity in a tropical montane forest. *Global Change Biology* 20: 3646–3659.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7: 335–336.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, USA* 108: 4516–4522.

Chaiyasen, A., J. P. W. Young, N. Teaumroong, P. Gavinlertvatana, and S. Lumyong. 2014. Characterization of arbuscular mycorrhizal fungus communities of *Aquilaria crassna* and *Tectona grandis* roots and soils in Thailand plantations. *PLoS ONE* 9: e112591.

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

Davison, J., M. Öpik, M. Zobel, M. Vasar, M. Metsis, and M. Moora. 2012. Communities of arbuscular mycorrhizal fungi detected in forest soil are spatially heterogeneous but do not vary throughout the growing season. *PLoS ONE* 7: e41938.

Doane, T. A., and W. R. Horwath. 2003. Spectrophotometric determination of nitrate with a single reagent. *Analytical Letters* 36: 2713–2722.

Dumbrell, A. J., M. Nelson, T. Helgason, C. Dytham, and A. H. Fitter. 2010. Idiosyncrasy and overdominance in the structure of natural communities of arbuscular mycorrhizal fungi: Is there a role for stochastic processes? *Journal of Ecology* 98: 419–428.

Dumbrell, A. J., P. D. Ashton, N. Aziz, G. Feng, M. Nelson, C. Dytham, A. H. Fitter, and T. Helgason. 2011. Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. *New Phytologist* 190: 794–804.

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics (Oxford, England)* 27: 2194–2200.

Egerton-Warburton, L. M., N. Collins Johnson, and E. B. Allen. 2007. Mycorrhizal community dynamics following nitrogen fertilizations: A cross-site test in five grasslands. *Ecological Monographs* 77: 527–544.

Eom, A., D. C. Hartnett, and G. W. T. Wilson. 2000. Host plant species effects on arbuscular mycorrhizal fungal communities in tallgrass prairie. *Oecologia* 122: 435–444.

Hassan, S. E. D., E. Boon, M. St-Arnaud, and M. Hijri. 2011. Molecular biodiversity of arbuscular mycorrhizal fungi in trace metal-polluted soils. *Molecular Ecology* 20: 3469–3483.

House, G. L., S. Ekanayake, Y. Ruan, U. M. E. Schütte, W. Kaonongbua, G. Fox, Y. Ye, and J. D. Bever. 2016. Phylogenetically structured differences in rRNA gene sequence variation among species of arbuscular mycorrhizal fungi and their implications for sequence clustering. *Applied and Environmental Microbiology* 82: 4921–4930.

Labidi, S., F. B. Jeddi, B. Tisserant, D. Debiane, S. Rezgui, A. Grandmougin-Ferjani, and A. L. H. Sahraoui. 2012. Role of arbuscular mycorrhizal symbiosis in root mineral uptake under $CaCO_3$ stress. *Mycorrhiza* 22: 337–345.

Lee, J., S. Lee, and J. P. W. Young. 2008. Improved PCR primers for the detection and identification of arbuscular mycorrhizal fungi. *FEMS Microbiology Ecology* 65: 339–349.

Lekberg, Y., R. T. Koide, J. R. Rohr, L. Aldrich-Wolfe, and J. B. Morton. 2007. Role of niche restrictions and dispersal in the composition of arbuscular mycorrhizal fungal communities. *Journal of Ecology* 95: 95–105.

López-Martínez, J. O., L. Sanaphre-Villanueva, J. M. Dupuy, J. L. Hernández-Stefanoni, J. A. Meave, and J. A. Gallardo-Cruz. 2013. β-Diversity of functional groups of woody plants in a tropical dry forest in Yucatan. *PLoS ONE* 8: e73660.

Oksanen, J., F. Guillame Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, et al. 2015. vegan: Community ecology package. R package version 2.2-1. Website https://cran.r-project.org/web/packages/vegan/index.html [accessed 2 August 2017].

Öpik, M., M. Metsis, T. J. Daniell, M. Zobel, and M. Moora. 2009. Large-scale parallel 454 sequencing reveals host ecological group specificity of arbuscular mycorrhizal fungi in a boreonemoral forest. *New Phytologist* 184: 424–437.

Öpik, M., A. Vanatoa, E. Vanatoa, M. Moora, J. Davison, J. M. Kalwij, Ü. Reier, and M. Zobel. 2010. The online database Maarj*AM* reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytologist* 188: 223–241.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Redecker, D., A. Schüßler, H. Stockinger, S. L. Stürmer, J. B. Morton, and C. Walker. 2013. An evidence-based consensus for the classification of arbuscular mycorrhizal fungi (Glomeromycota). *Mycorrhiza* 23: 515–531.

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

Sanders, I. R., and A. Rodriguez. 2016. Aligning molecular studies of mycorrhizal fungal diversity with ecologically important levels of diversity in ecosystems. *ISME Journal* 10: 2780–2786.

Schüßler, A., and C. Walker. 2010. The Glomeromycota. A species list with new families and new genera. Website http://www.amf-phylogeny.com [accessed 28 July 2017].

Simon, L., M. Lalonde, and T. D. Bruns. 1992. Specific amplification of 18S fungal ribosomal genes from VA endomycorrhizal fungi colonizing roots. *Applied and Environmental Microbiology* 58: 291–295.

Smith, D. P., and K. G. Peay. 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS ONE* 9: e90234.

van der Heijden, M. G. A., T. Boller, A. Wiemken, and I. R. Sanders. 1998. Different arbuscular mycorrhizal fungal species are potential determinants of plant community structure. *Ecology* 79: 2082–2091.

Walters, W. A., J. Caporaso, C. L. Lauber, D. Berg-Lyons, N. Fierer, and R. Knight. 2011. PrimerProspecter: de novo design and taxonomic analysis of PCR primers. *Bioinformatics (Oxford, England)* 27: 1159–1161.

Wang, C., P. J. White, and L. Chunjian. 2016. Colonization and community structure of arbuscular mycorrhizal fungi in maize roots at different depths in the soil profile respond differently to phosphorus inputs on a long-term experimental site. *Mycorrhiza* 27: 1–13.

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73: 5261–5267.

Weatherburn, M. W. 1967. Phenol-hypochlorite reaction for determination of ammonia. *Analytical Chemistry* 39: 971–974.

Wilson, H., B. R. Johnson, B. Bohannan, L. Pfeifer-Meister, R. Mueller, and S. D. Bridgham. 2016. Experimental warming decreases arbuscular mycorrhizal fungal colonization in prairie plants along a Mediterranean climate gradient. *PeerJ* 4: e2083.

Appendix 1. Levels of soil nitrate (NO_3), ammonium (NH_4), phosphate (P), pH, and electrical conductivity (EC) in soil samples from each study site. Data are presented as site means with standard deviation in parentheses.*

Site (No. samples)	NO_3 (μg g soil^{-1})	NH_4 (μg g soil^{-1})	P (μg g soil^{-1})	pH	EC (μS cm^{-1})
El Eden (n = 14)	3 (3)[a]	12 (8)[a]	13 (5)[a]	7.62 (0.6)[a]	146 (41)[a]
La Higuera (n = 19)	3 (2)[a]	32 (11)[b]	21 (11)[b]	8.34 (0.3)[b]	173 (60)[a]

*Within each column, means with the same letter do not differ significant at $P < 0.05$ (Tukey's HSD test).

Appendix 2. Twelve-base Golay error-correcting barcode sequences that passed in silico testing for secondary structure formation. Individual barcode sequences from this list are substituted for XXXXXXXXXXXX in the "Golay barcode" region (Table 2) to generate indexed reverse primer constructs for PCR.

Barcode name	Barcode nucleotide sequence	Barcode name	Barcode nucleotide sequence	Barcode name	Barcode nucleotide sequence
AML2_il001	TCCCTTGTCTCC	AML2_il027	AGTTACGAGCTA	AML2_il053	CGGTCAATTGAC
AML2_il002	GCTGTACGGATT	AML2_il028	GCATATGCACTG	AML2_il054	GTGGAGTCTCAT
AML2_il003	ATCACCAGGTGT	AML2_il029	CAACTCCCGTGA	AML2_il055	GCTCGAAGATTC
AML2_il004	TGGTCAACGATA	AML2_il030	TTGCGTTAGCAG	AML2_il056	AGGCTTACGTGT
AML2_il005	ATCGCACAGTAA	AML2_il031	TACGAGCCCTAA	AML2_il057	TCTCTACCACTC
AML2_il006	AGCGGAGGTTAG	AML2_il032	CACTACGCTAGA	AML2_il058	ACTTCCAACTTC
AML2_il007	ATCCTTTGGTTC	AML2_il033	TGCAGTCCTCGA	AML2_il059	CTCACCTAGGAA
AML2_il008	TACAGCGCATAC	AML2_il034	ACCATAGCTCCG	AML2_il060	GTGTTGTCGTGC
AML2_il009	ACCGGTATGTAC	AML2_il035	TCGACATCTCTT	AML2_il061	CCACAGATCGAT
AML2_il010	AATTGTGTCGGA	AML2_il036	GAACACTTTGGA	AML2_il062	TATCGACACAAG
AML2_il011	TGCATACACTGG	AML2_il037	GAGCCATCTGTA	AML2_il063	GATTCCGGCTCA
AML2_il012	AGTCGAACGAGG	AML2_il038	TAATACGGATCG	AML2_il064	TAGGCATGCTTG
AML2_il013	ACCAGTGACTCA	AML2_il039	TCGGAATTAGAC	AML2_il065	AACTAGTTCAGG
AML2_il014	GAATACCAAGTC	AML2_il040	TGTGAATTCGGA	AML2_il066	GTACGATATGAC
AML2_il015	GTAGATCGTGTA	AML2_il041	TACTACGTGGCC	AML2_il067	TAGTATGCGCAA
AML2_il016	CCAATACGCCTG	AML2_il042	GGCCAGTTCCTA	AML2_il068	ATGGCTGTCAGT
AML2_il017	GATCTGCGATCC	AML2_il043	GATGTTCGCTAG	AML2_il069	GCGTTCTAGCTG
AML2_il018	CAGCTCATCAGC	AML2_il044	CTATCTCCTGTC	AML2_il070	GTTGTTCTGGGA
AML2_il019	CAAACAACAGCT	AML2_il045	ACTCACAGGAAT	AML2_il071	ATGTCACCGCTG
AML2_il020	GCAACACCATCC	AML2_il046	ATGATGAGCCTC	AML2_il072	AGCAGAACATCT
AML2_il021	CGAGCAATCCTA	AML2_il047	GTCGACAGAGGA	AML2_il073	TGGAGTAGGTGG
AML2_il022	AGTCGTGCACAT	AML2_il048	TGTCGCAAATAG	AML2_il074	TTGGCTCTATTC
AML2_il023	GTATCTGCGCGT	AML2_il049	CATCCCTCTACT	AML2_il075	GATCCCACGTAC
AML2_il024	CGAGGGAAAGTC	AML2_il050	TATACCGCTGCG	AML2_il076	TACCGCTTCTTC
AML2_il025	CAAATTCGGGAT	AML2_il051	AGTTGAGGCATT	AML2_il077	TGTGCGATAACA
AML2_il026	AGATTGACCAAC	AML2_il052	ACAATAGACACC	AML2_il078	GATTATCGACGA
				AML2_il079	GCCTAGCCCAAT

Appendix 3. Annotated QIIME workflow.

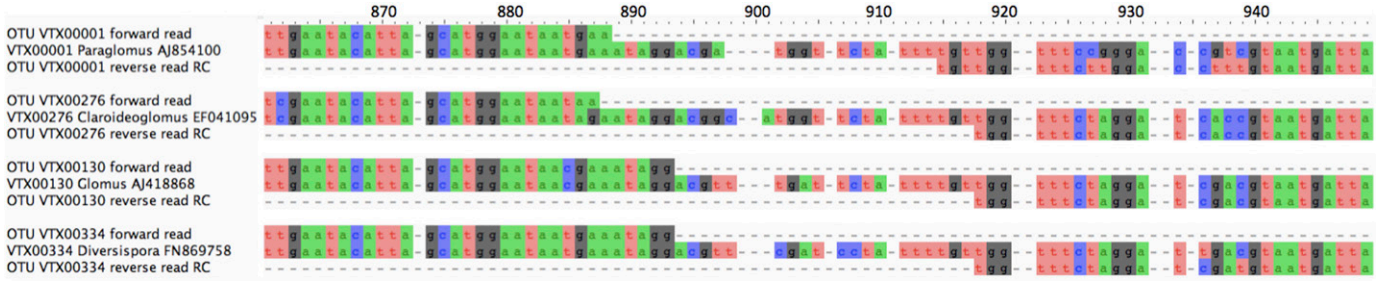
```
###################################
# Front matter
###################################
# Annotated QIIME workflow
# Performed with QIIME v1.9.1
# Uses four MaarjAM database resource files (Öpik M, personal communication)
## MaarjAM Virtual Taxon type sequences: maarjam_vt_types_from_31-03-2013.fasta
## Taxonomy strings for VT type sequences: vt_types_tax_from_31-03-2013.txt
## MaarjAM complete database: maarjAM.5.fasta
## Taxonomy strings for complete database: maarjAM.id_to_taxonomy.5.txt

###################################
# Demultiplexing and quality filtering raw sequencing reads
###################################
## -i path to input data in fastq format: used raw read 1 sequence data output by Miseq
## -b path to index read data in fastq format: used raw index read data output by Miseq
## -m path to a tab-delimited text file containing barcode sequences for each sampe ID
## -o path to directory where demultiplexing output will be generated
## -q sets the maximum unacceptable quality score: used 19 (only qualities >20 accepted)
##–barcode_type sets number of bases in barcode: used golay_12 to indicated use of
## 12 base Golay error correcting barcodes
##–rev_comp_mapping_barcodes set to use reverse complement of barcode sequences
##–max_barcode_errors sets maximum number of errors in index read to still allow
## assignment to a sample: used 2.5, so index reads with three or more errors remain
## unassigned

split_libraries_fastq.py -i /PATH/TO/raw_data_fwd_reads.fastq -b /PATH/TO/raw_data_index_reads.fastq -m /PATH/TO/demultiplex_map.txt -o /PATH/TO/demultiplexing_output/ -q 19–barcode_type golay_12–rev_comp_mapping_barcodes–max_barcode_errors 2.5

### NOTE: This step processes raw sequencing output for 59 samples that were
### multiplexed in a single run.
### Of the total samples, 11 had fewer than 1000 total quality filtered sequencing
### reads, and were excluded from all further sequence data processing and analysis.

###################################
# Removing samples with fewer than 1000 quality filtered sequencing reads
###################################
## -f path to input sequences in fasta format: used quality filtered sequences generated
## in previous step
## -o path to an output file in fasta format
##–sample_id_fp path to newline-delimited text file where each line contains one
## sample ID to discard: used a list of the 11 samples with fewer than 1000 reads
## -n sets to discard sequences indicated by–sample_id_fp
filter_fasta.py -f /PATH/TO/demultiplexing_output/seqs.fna -o /PATH/TO/demultiplexing_output/fwd_qc20_more100.fasta–sample_id_fp /PATH/TO/discard_ids.txt -n

###################################
# All subsequent steps were conducted three times, once for each OTU clustering threshold
###################################

###################################
# Clustering operational taxonomic units using an open reference OTU picking strategy
###################################
## -i path to input fasta sequence file: using the output of the previous step
## -r path to a reference database of fasta formatted sequences used for open reference
## otu picking: using the MaarjAM VT database
## -o path to the output directory where the biom formatted OTU table will be written
## -p path to a parameters file with appropriate OTU clustering threshold set
##–suppress_taxonomy_assignment prevents taxonomy assignment at this step
##–suppress_align_and_tree prevents OTU alignment and tree building at this step
##–min_otu_size sets the minimum number of constituent sequences in an OTU for it to be
## retained in the final OTU table: set to 1 to include all OTUs

pick_open_reference_otus.py -i /PATH/TO/demultiplexing_output/fwd_qc20_more100.fasta -r /PATH/TO/maarjAM_vt_types_from_31-03-2013.fasta -o /PATH/TO/otu_table_output/ -p /PATH/TO/params.txt–suppress_taxonomy_assignment–suppress_align_and_tree–min_otu_size 1

###################################
# Excluding probable artifacts and non-target taxa using BLAST stand-alone command line
# tools Nucleotide-Nucleotide BLAST 2.2.26+ (Altschul et al., 1990) against the MaarjAM.5
# database of reference AMF SSU sequences.
###################################

## -query path to representative sequences generated by OTU picking step: using rep_set.fna
## generated by the previous step
```

```
## -db path to the database used for BLAST alignments: using the MaarjAM complete database
## -out path to the output directory where alignment table will be stored
## -evalue minimum acceptable BLAST expect value score for a representative sequences to
## be retained for further analysis: set to 1e-50
## -perc_identity sets minimum required percent sequence identity across aligned region for a hit
## to be recorded: set to same level as OTU clustering similarity threshold.
## -qcov_hsp_perc sets the minimum percent of the input sequence length that must align to a
## reference sequence to record a hit: set at 90%
## - num_alignments sets the total number of alignments to record: set to 1
## -outfmt sets output formatting options: set to 6 (tabular output)

blastn -query /PATH/TO/otu_table_output/rep_set.fna -db /PATH/TO/maarjAM5blastDB -out /PATH/TO/blast_maarjam_output/blast_table_maarjam.txt -evalue 1e-50 -perc_identity 90 -qcov_hsp_perc 90 -num_alignments 1 -outfmt "6 qseqid sseqid pident qcovhsp length mismatch gapopen qstart qend sstart send evalue bitscore" -num_threads 2

#################################
# Making fasta files for OTUs with and without hits to the MaarjAM database
#################################
## filter_fasta.py parameters as above

awk '{print $1}' /PATH/TO/blast_maarjam_output/blast_table_maarjam.txt > /PATH/TO/blast_maarjam_output/match_ids.txt
filter_fasta.py -f /PATH/TO/otu_table_output/rep_set.fna -s /PATH/TO/blast_maarjam_output/match_ids.txt -o PATH/TO/blast_maarjam_output/matching.fna
filter_fasta.py -f /PATH/TO/otu_table_output/rep_set.fna -s /PATH/TO/blast_maarjam_output/match_ids.txt -o PATH/TO/blast_maarjam_output/non-matching.fna -n

#################################
# Excluding primer artifacts using BLAST implemented in QIIME
#################################
## -i path to representative sequences generated by OTU picking step: using rep_set.fna
## generated by the previous step
## -d path to a database of PCR primer construct sequences, available in Table 1 and Appendix 1
## -o path to the output directory where results will be stored
## -e minimum acceptable BLAST expect value score for a representative sequences to
## be retained for further analysis: set to default 1e-10

exclude_seqs_by_blast.py -i /PATH/TO/blast_maarjam_output/matching.fna -d /PATH/TO/pcr_primer_constructs.fasta -o /PATH/TO/primer_artifact_exclusion_output/ -p 0.5

#################################
# Excluding chimeras using uCHIME v4.2 (Edgar et al., 2011) in database mode
#################################
##–input path to OTU sequences that passed previous BLAST screenings
##–db path to a database of all MaarjAM.5 forward and reverse complement sequences
##–uchimeout path to a text file to store uCHIME result scores
## -uchimealns path to a text file to store uCHIME alignments
uchime–input /PATH/TO/primer_artifact_exclusion_output/non-matching.fna–db /PATH/TO/full_maarjAM_fwd_and_rc.fasta–uchimeout /PATH/TO/uchime_results/results.uchime–uchimealns /PATH/TOuchime_results/results.alns

#################################
# Excluding non-target taxa using BLAST stand-alone command line tools
# Nucleotide-Nucleotide BLAST 2.2.26+ (Altschul et al., 1990) against the NCBI non-redundant
# nucleotide database.
#################################
## -query path to OTU sequences that passed previous BLAST screenings
## -db path to a locally stored version of the NCBI nt database
## -out path to a text file to store output
## other parameters as above

blastn -query /PATH/TO/primer_artifact_exclusion_output/non-matching.fna -db nt -out /PATH/TO/blast_nt_output/blast_table_nt.txt -num_alignments 1 -outfmt "6 qseqid sseqid pident qcovhsp length mismatch gapopen qstart qend sstart send evalue bitscore" -num_threads 2

#################################
# Extracting hit taxonomic information from the nt database using BLAST stand-alone command
## line tools BLAST 2.2.26+ (Altschul et al., 1990)
#################################
## -entry_batch path to a file of subject IDs that hit OTU sequences
## -db path to a locally stored version of the NCBI nt database
## -out path to a text file to store output
## -outfmt sets fields to output: set to GI number and scientific name
blastdbcmd -entry_batch /PATH/TO/blast_nt_output/subject_ids.txt -db nt -out /PATH/TO/blast_nt_output/subject_names.txt -outfmt "%g %S"
#################################
# Removing artifactual, chimeric, and non-target reads from the set of OTU representative
# sequences
#################################
## -f path to original representative OTU sequences fasta file
## -s path to list of all OTUs that failed to pass blast and uCHIME screening
```

```
## -o path to a new representative sequence fasta file excluding screened OTUs
## -n as above

filter_fasta.py -f /PATH/TO/otus_table_output/rep_set.fna -s /PATH/TO/otus_to_exclude_blast.txt -o /PATH/TO/otu_table_output/screened_rep_set.fna -n

###################################
# Assigning taxonomy to screened OTUs using the QIIME implementation of RDP Classifier
###################################
## -i path to fasta formatted sequence file to assign taxonomies to: using representative ## sequences with acceptable BLAST hits from the previous step, matching.fna
## -r path to fasta formatted referece database that will be used to assign taxonomies to input
## sequences: using the MaarjAM.5 complete database
## -t path to taxonomy strings for reference database: using taxonomy file for complete database
## -o path to directory where taxonomic assignment files will be generated
## -m method of taxonomy assignment to use: set to use RDP Classifier
##–rdp_max_memory sets limit in KB RDP taxonomic assignment can use: set to 20000

assign_taxonomy.py -i / PATH/TO/otu_table_output/screened_rep_set.fna -r /PATH/TO/maarjAM.5.fasta -t /PATH/TO/maarjAM.id_to_taxonomy.5.txt -o /PATH/TO/
taxonomy_assignment_output/ -m rdp–rdp_max_memory 20000

###################################
# Making a new OTU table that excludes probable artifacts and non-target taxa and
# includes taxonomic assignments
###################################
## -i path to the final OTU map generated by initial OTU picking step
## -t path to taxonomic assignments generated in the previous step
## -e path to a list of all OTUs that were excluded from taxonomy assignment
## -o path to new biom formatted OTU table that will be generated

make_otu_table.py -i /PATH/TO/otu_table_output/final_otu_map.txt -t /PATH/TO/taxonomy_assignment_output/matching_tax_assignments.txt -e /PATH/TO/otus_
to_exclude.txt -o /PATH/TO/otu_table_output/screened_rdptax_otu_table.biom

###################################
# Creating the All OTUs OTU table for this clustering threshold
###################################
## -i path to an input biom formatted OTU table: using the output of the previous step
## -o path to generate a new biom formatted OTU table with information about a subset
## of samples from the input file
##–sample_id_fp path to a text file identifying samples to be discarded in the new
## OTU table: set to restrict the OTU table to the 33 samples used in this study
##–negate_sample_id_fp sets list of files to be discarded rather than kept

filter_samples_from_otu_table.py -i /PATH/TO/otu_table_output/screened_rdptax_otu_table.biom -o /PATH/TO/otu_table_output/otu_table_All.biom–sample_id_fp
/PATH/TO/samples_to_exclude.txt –negate_sample_id_fp

###################################
# Creating the 2+ton OTUs OTU table for this clustering threshold
###################################
## -i path to an input biom formatted OTU table: using the All OTUs table
## -o path to generate a new biom formatted OTU table with information about a subset
## of samples from the input file
## -n minimum read abundance for OTU to be kept: set to 2

filter_otus_from_otu_table.py -i /PATH/TO otu_table_output/otu_table_All.biom -o /PATH/TO/otu_table_output/otu_table_mc2.biom -n 2

###################################
# Creating the 10+ton OTUs OTU table for this clustering threshold
###################################
## -i path to an input biom formatted OTU table: using the All OTUs table
## -o path to generate a new biom formatted OTU table with information about a subset
## of samples from the input file
## -n minimum read abundance for OTU to be kept: set to 10

filter_otus_from_otu_table.py -i /PATH/TO otu_table_output/otu_table_All.biom -o /PATH/TO/otu_table_output/otu_table_mc10.biom -n 10

###################################
# For each OTU table
###################################

###################################
# Making alpha-rarefaction plots
###################################
## -i path to input biom formatted OTU table: using output of previous step
## -m path to a tab-delimited text map file with sample metadata
## -o path to directory where alpha rarefaction output will be generated
## -p path to parameters file setting indices to use
```

```
alpha_rarefaction.py -i /PATH/TO/otu_table_output/otu_table_All.biom -m /PATH/TO/map.txt -o /PATH/TO/otu_table_All_alpha_rarefaction_output -p /PATH/TO/params.txt

###################################
# Generating evenly sampled beta-diversity distance matrices, principal coordinates,
# and 3D principal coordinates ordination plots
###################################
## -i path to input biom formatted OTU table: uses eleven_samples table generated above
## -m path to a tab delimited text map file with sample metadata
## -p path to parameters file setting diversity indices to use
## -o path to directory where beta diversity output will be generated
## -e sets number of sequences to subsample: set to 1000

beta_diversity_through_plots.py -i /PATH/TO/otu_table_output/otu_table_All.biom -m /PATH/TO/map.txt -p /PATH/TO/params.txt -o /PATH/TO/otu_table_All_beta_diversity_output/ -e 1000
```

Appendix 4. Example alignments of forward and reverse reads assigned to Virtual Taxon sequences from the Maarj*AM* database showing approximately 30 base regions separating reads and preventing successful assembly. Numbers represent position in MAFFT alignment of Maarj*AM* Virtual Taxon database to which the sequence reads were aligned.



Appendix 5. Taxonomic information for the arbuscular mycorrhizal fungal virtual taxa detected in soil and root samples and their respective morphologically described species, where available.[a]

GenBank accession no.	Virtual taxa file names	Family	Genus	Species
EF041095	VTX00276	Claroideoglomeraceae	*Claroideoglomus*	None
AJ315524	VTX00054	Diversisporaceae	*Diversispora*	*Diversispora celata* C. Walker, Gamper & A. Schüßler; *D. auratia* (Blaszk., Blanke, Renker & Buscot) C. Walker & A. Schüßler; *Entrophospora nevadensis* Palenz., N. Ferrol, Azcón-Aguilar, & Oehl
AM849296	VTX00060	Diversisporaceae	*Diversispora*	*Diversispora celata*; *D. auratia*; *D. eburnea* (L. J. Kenn., J. C. Stutz & J. B. Morton) C. Walker & A. Schüßler; *Entrophospora nevadensis*
X86687	VTX00061	Diversisporaceae	*Diversispora*	*Diversispora epigaea* (B. A. Daniels & Trappe) C. Walker & A. Schüßler; *D. spurca* (C. M. Pfeiff., C. Walker & Bloss) C. Walker & A. Schüßler; *D. versiformis* (P. Karst.) Oehl, G. A. Silva & Sieverd.
Y17650	VTX00263	Diversisporaceae	*Diversispora*	*Diversispora spurca*
FR686957	VTX00347	Diversisporaceae	*Diversispora*	*Diversispora trimurales* (Koske & Halvorson) C. Walker & A. Schüßler
EU332707	VTX00040	Diversisporaceae	*Diversispora*	None
AY129577	VTX00059	Diversisporaceae	*Diversispora*	None
FN869704	VTX00380	Diversisporaceae	*Diversispora*	None
U96146	VTX00039	Gigasporaceae	*Scutellospora*	*Gigaspora decipiens* I. R. Hall & L. K. Abbott
AJ306436	VTX00254	Gigasporaceae	*Scutellospora*	*Scutellospora spinosissima* C. Walker & Cuenca
AJ418851	VTX00041	Gigasporaceae	*Scutellospora*	*Scutellospora weresubiae* Koske & C. Walker; *Racocetra castanea* (C. Walker) Oehl, F. A. Souza & Sieverd.; *R. gregaria* (N. C. Schenck & T. H. Nicolson) Oehl, F. A. Souza & Sieverd.; *R. fulgida* (Koske & C. Walker) Oehl, F. A. Souza & Sieverd.; *R. persica* (Koske & C. Walker) Oehl, F. A. Souza & Sieverd.
AJ306434	VTX00255	Gigasporaceae	*Scutellospora*	*Dentiscutata heterogama* (T. H. Nicolson & Gerdemann) C. Walker & F. E. Sanders; *Dentiscutata reticulata* (Koske, D. D. Mill. & C. Walker) Sieverd., F. A. Souza & Oehl.
AJ315526	VTX00104	Glomeraceae	*Glomus*	*Glomus* cf. *microaggregatum* Koske, Gemma & P. D. Olexia
DQ164825	VTX00155	Glomeraceae	*Glomus*	*Glomus iranicum* Blaszk., Kovács & Balázs
AM849311	VTX00199	Glomeraceae	*Glomus*	*Glomus macrocarpum* Tul. & C. Tul.; *G. hoi* S. M. Berch & Trappe
FJ164237	VTX00287	Glomeraceae	*Glomus*	*Glomus perpusillum* Blaszk. & Kovács
AF213462	VTX00099	Glomeraceae	*Glomus*	*Glomus proliferum* Dalpé & Declerck
AY129592	VTX00069	Glomeraceae	*Glomus*	*Glomus sinuosum* (Gerd. & B. K. Bakshi) R. T. Almeida & N. C. Schenck
AJ496056	VTX00115	Glomeraceae	*Glomus*	*Glomus vesiculiferum* (Thaxt.) Gerd. & Trappe
AJ505617	VTX00105	Glomeraceae	*Glomus*	*Rhizophagus intraradices* (N. C. Schenck & G. S. Sm.) C. Walker & A. Schüßler
AJ418876	VTX00113	Glomeraceae	*Glomus*	*Rhizophagus intraradices*
AM849267	VTX00114	Glomeraceae	*Glomus*	*Rhizophagus intraradices*; *R. irregularis* (Blaszkowski, Wubet, Renker & Buscot) C. Walker & A. Schüßler
AM849308	VTX00064	Glomeraceae	*Glomus*	*Septoglomus furcatum* Blaszk., Chwat, Kovács & Ryszka; *S. constrictum* (Trappe) Sieverd., G. A. Silva & Oehl; *S. xanthium* (Blaszk., Blanke, Renker & Buscot) G. A. Silva, Oehl & Sieverd.; *Glomus africanum* Blaszk. & Kovács
AJ505812	VTX00063	Glomeraceae	*Glomus*	*Septoglomus viscosum* (T. H. Nicolson) C. Walker, D. Redecker, D. Stiller & A. Schüßler; *Diversispora* sp.
DQ336485	VTX00053	Glomeraceae	*Glomus*	None
DQ336448	VTX00068	Glomeraceae	*Glomus*	None
DQ371669	VTX00076	Glomeraceae	*Glomus*	None
AB365818	VTX00077	Glomeraceae	*Glomus*	None
AY129614	VTX00078	Glomeraceae	*Glomus*	None
AB183952	VTX00084	Glomeraceae	*Glomus*	None
AB365850	VTX00085	Glomeraceae	*Glomus*	None
AJ563892	VTX00086	Glomeraceae	*Glomus*	None
AY129635	VTX00087	Glomeraceae	*Glomus*	None
AY512364	VTX00089	Glomeraceae	*Glomus*	None
DQ336444	VTX00091	Glomeraceae	*Glomus*	None
AB365822	VTX00092	Glomeraceae	*Glomus*	None
EU332715	VTX00093	Glomeraceae	*Glomus*	None
AY129604	VTX00096	Glomeraceae	*Glomus*	None
AB326008	VTX00100	Glomeraceae	*Glomus*	None

Appendix 5. Continued.

GenBank accession no.	Virtual taxa file names	Family	Genus		Species
EU350053	VTX00103	Glomeraceae	*Glomus*	None	
EU350068	VTX00107	Glomeraceae	*Glomus*	None	
AY330278	VTX00108	Glomeraceae	*Glomus*	None	
AY129575	VTX00109	Glomeraceae	*Glomus*	None	
EU417585	VTX00111	Glomeraceae	*Glomus*	None	
DQ336482	VTX00112	Glomeraceae	*Glomus*	None	
DQ396700	VTX00117	Glomeraceae	*Glomus*	None	
AF437667	VTX00120	Glomeraceae	*Glomus*	None	
AF437663	VTX00121	Glomeraceae	*Glomus*	None	
AF480153	VTX00123	Glomeraceae	*Glomus*	None	
AM849263	VTX00125	Glomeraceae	*Glomus*	None	
AY129611	VTX00126	Glomeraceae	*Glomus*	None	
DQ085198	VTX00128	Glomeraceae	*Glomus*	None	
AJ418868	VTX00130	Glomeraceae	*Glomus*	None	
AB365855	VTX00131	Glomeraceae	*Glomus*	None	
AY129605	VTX00132	Glomeraceae	*Glomus*	None	
AJ563890	VTX00137	Glomeraceae	*Glomus*	None	
AJ563896	VTX00140	Glomeraceae	*Glomus*	None	
AB365857	VTX00146	Glomeraceae	*Glomus*	None	
DQ396751	VTX00154	Glomeraceae	*Glomus*	None	
AJ563861	VTX00156	Glomeraceae	*Glomus*	None	
AM849314	VTX00160	Glomeraceae	*Glomus*	None	
AM849298	VTX00163	Glomeraceae	*Glomus*	None	
EF154349	VTX00165	Glomeraceae	*Glomus*	None	
AJ418860	VTX00166	Glomeraceae	*Glomus*	None	
EU350060	VTX00167	Glomeraceae	*Glomus*	None	
AM412105	VTX00175	Glomeraceae	*Glomus*	None	
DQ336480	VTX00183	Glomeraceae	*Glomus*	None	
AB365808	VTX00185	Glomeraceae	*Glomus*	None	
AM412533	VTX00186	Glomeraceae	*Glomus*	None	
AM849326	VTX00187	Glomeraceae	*Glomus*	None	
AM849257	VTX00194	Glomeraceae	*Glomus*	None	
AM746134	VTX00197	Glomeraceae	*Glomus*	None	
AJ563889	VTX00202	Glomeraceae	*Glomus*	None	
DQ371690	VTX00204	Glomeraceae	*Glomus*	None	
AY129588	VTX00206	Glomeraceae	*Glomus*	None	
AY129586	VTX00209	Glomeraceae	*Glomus*	None	
AJ699061	VTX00213	Glomeraceae	*Glomus*	None	
AF074370	VTX00214	Glomeraceae	*Glomus*	None	
DQ336508	VTX00217	Glomeraceae	*Glomus*	None	
AB183976	VTX00224	Glomeraceae	*Glomus*	None	
AJ496098	VTX00233	Glomeraceae	*Glomus*	None	
DQ357117	VTX00234	Glomeraceae	*Glomus*	None	
AB183981	VTX00246	Glomeraceae	*Glomus*	None	
AY129627	VTX00247	Glomeraceae	*Glomus*	None	
AB365803	VTX00248	Glomeraceae	*Glomus*	None	
AM746148	VTX00253	Glomeraceae	*Glomus*	None	
EU332734	VTX00256	Glomeraceae	*Glomus*	None	
DQ371674	VTX00269	Glomeraceae	*Glomus*	None	
AB365831	VTX00270	Glomeraceae	*Glomus*	None	
AM412083	VTX00280	Glomeraceae	*Glomus*	None	
DQ396779	VTX00293	Glomeraceae	*Glomus*	None	
EF154586	VTX00294	Glomeraceae	*Glomus*	None	
EU169401	VTX00302	Glomeraceae	*Glomus*	None	

APPENDIX 5. Continued.

GenBank accession no.	Virtual taxa file names	Family	Genus		Species
FM875902	VTX00304	Glomeraceae	*Glomus*	None	
FM876953	VTX00311	Glomeraceae	*Glomus*	None	
FN263137	VTX00312	Glomeraceae	*Glomus*	None	
EU340294	VTX00322	Glomeraceae	*Glomus*	None	
GU183691	VTX00323	Glomeraceae	*Glomus*	None	
EU340316	VTX00326	Glomeraceae	*Glomus*	None	
GU353949	VTX00331	Glomeraceae	*Glomus*	None	
FN859983	VTX00333	Glomeraceae	*Glomus*	None	
FN869758	VTX00334	Glomeraceae	*Glomus*	None	
FN429114	VTX00342	Glomeraceae	*Glomus*	None	
FN556624	VTX00344	Glomeraceae	*Glomus*	None	
HF566497	VTX00360	Glomeraceae	*Glomus*	None	
FR821540	VTX00363	Glomeraceae	*Glomus*	None	
HF566504	VTX00364	Glomeraceae	*Glomus*	None	
HF566507	VTX00372	Glomeraceae	*Glomus*	None	
HE798788	VTX00382	Glomeraceae	*Glomus*	None	
HF566487	VTX00397	Glomeraceae	*Glomus*	None	
HE798777	VTX00398	Glomeraceae	*Glomus*	None	
HE798804	VTX00399	Glomeraceae	*Glomus*	None	
HE615074	VTX00409	Glomeraceae	*Glomus*	None	
GQ140605	VTX00412	Glomeraceae	*Glomus*	None	
FM955473	VTX00416	Glomeraceae	*Glomus*	None	
AY330274	VTX00417	Glomeraceae	*Glomus*	None	
FN869759	VTX00418	Glomeraceae	*Glomus*	None	
FN646035	VTX00335	Paraglomerales	*Paraglomus*	*Paraglomus majewskii* Blaszk.	
AJ854100	VTX00001	Paraglomerales	*Paraglomus*	None	

[a] Accession numbers from GenBank, virtual taxa accession names, and family and genus names are taken from Maarj*AM* (accessed 9 June 2017; http://maarjam.botany.ut.ee); morphological species nomenclature follows Redecker et al. (2013).