BRIEF REPORT

An updated LSU database and pipeline for environmental DNA identification of arbuscular mycorrhizal fungi

Camille S. Delavaux[1] · Robert J. Ramos[2] · Sidney L. Stürmer[3] · James D. Bever[4,5]

Abstract

Recent work established a backbone reference tree and phylogenetic placement pipeline for identification of arbuscular mycorrhizal fungal (AMF) large subunit (LSU) rDNA environmental sequences. Our previously published pipeline allowed any environmental sequence to be identified as putative AMF or within one of the major families. Despite this contribution, difficulties in implementation of the pipeline remain. Here, we present an updated database and pipeline with (1) an expanded backbone tree to include four newly described genera and (2) several changes to improve ease and consistency of implementation. In particular, packages required for the pipeline are now installed as a single folder (conda environment) and the pipeline has been tested across three university computing clusters. This updated backbone tree and pipeline will enable broadened adoption by the community, advancing our understanding of these ubiquitous and ecologically important fungi.

Camille S. Delavaux and Robert J. Ramos contributed and share co-first authorship

✉ Camille S. Delavaux
camille.delavaux@usys.ethz.ch

1   Institute of Integrative Biology, ETH Zurich (Swiss Federal Institute of Technology), Universitätsstrasse 16, Zurich 8092, Switzerland

2   The Environmental Data Science Innovation & Inclusion Lab (ESIIL), University of Colorado Boulder, Colorado 80309, USA

3   Departamento de Ciências Naturais, Universidade Regional de Blumenau, Blumenau, Santa Catarina 89030-903, Brazil

4   Department of Ecology and Evolutionary Biology, The University of Kansas, 2041 Haworth Hall, 1200 Sunnyside Avenue, Lawrence, KS 66045, USA

5   Kansas Biological Survey, The University of Kansas, 106 Higuchi Hall, 2101 Constant Avenue, Lawrence, KS 66047, USA

Introduction

The use of environmental DNA and amplicon sequencing has become a fundamental tool for fungal ecologists (Tedersoo et al. 2022), and particularly for those studying mycorrhizal fungi. These approaches enable researchers to obtain a relatively unbiased snapshot of the community using a very small amount of starting material (e.g., soil or roots). In theory, this should generate large amounts of data to understand the biogeography and ecology of this important group of fungi. However, the most common type of mycorrhizal fungi – the arbuscular mycorrhizal fungi (AMF, phylum Glomeromycota) – are not well represented when using the "universal" fungal primers targeting the internal transcribed spacer of the rDNA gene (Schoch et al. 2012; Lekberg et al. 2018; Tedersoo et al. 2022). Furthermore, AMF sequence databases are geographically biased and incomplete (Bidartondo et al., Öpik et al. 2010; Stockinger et al. 2010). Therefore, the full potential of this technology for AMF requires use of other gene regions and phylogenetic placement approaches.

Although there is no consensus on the ideal region to target or sequencing platforms to use for AMF, the large subunit (LSU) amplicon (short read) sequencing is a good

choice for next generation sequencing (Illumina; Stockinger et al. 2010; Delavaux et al. 2021, Delavaux et al. 2022). This region allows phylogenetic placement and subsequent discovery of novel taxa, while showing good taxonomic resolution at the species level (Krüger et al. 2009; Hart et al. 2015; House et al. 2016). Previous work established a curated reference tree and developed a pipeline for phylogenetic placement of AMF for the LSU region (Delavaux et al. 2021, 2022). Specifically, this work enabled phylogenetic placement of any environmental sequence into the backbone reference tree. Using these tools, environmental sequences can be placed into the Glomeromycota phylum – describing putative AMF – and further into the major 11 families. Importantly, the concerted effort packaging this process into a formal pipeline improved accessibility of this phylogenetic approach for interested scientists.

Despite the benefit to the research community brought by the new pipeline, users unfamiliar with command line interfaces and with low computing cluster support struggled with implementation, resulting in a substantial hurdle to the systematic adoption of this pipeline for those using the LSU. The main issues with implementation are trouble installing packages as required by the pipeline (specific versions and packages that require a conda environment, Anaconda 2024) and cluster specific assumptions in bash job scripts. Here, we update the database and pipeline by (1) expanding the backbone tree to include newly described genera and (2) further improve usability of this pipeline, including creating a conda requirements package hosting all required packages and testing the pipeline across three university clusters. This updated backbone reference tree and user-friendly pipeline will contribute to broadened adoption of this tool, ultimately improving the scientific community's understanding of the ecology of these important fungi (Table 1).

Expanded backbone reference tree

We expanded the backbone reference tree used in the pipeline (Delavaux et al. 2021; Delavaux et al. 2022) to include several new genera and species, reflecting changes occurring since the reference tree was first created (Fig. 1, version 16). In particular, we add species from the four genera

Epigeocarpum, *Silvaspora*, *Complexispora*, and *Entrophospora* (Błaszkowski et al. 2021a, b, 2022, 2023; da Silva et al. 2023; NCBI accession MW507157.1, MW541060, OQ437298, OQ437305, OQ437315.1, ON950380.1, ON950390.1). We used the start of the LSU region using primers LROR/FLR2 (Bunyard et al. 1994; Trouvelot et al. 1999; ~700–900 bp) and built our backbone tree using RAxML 8.2.12 (Stamatakis 2014) with 1000 bootstrap replicates and the evolutionary model GTRGAMMA. In parallel, we have updated all pipeline scripts to reflect this novel tree. Future work could expand the current backbone tree further by allowing for phylogenetic extraction of genera and even species. Further improvements will be integrated into our Github repository (https://github.com/c383d893/AMF-LSU-Database-and-Pipeline2) as new taxa are described.

One installation for the entire pipeline

Although previous formalizing of the database and pipeline helped improve accessibility to phylogenetic placement for the LSU region, even this streamlined version may be intimidating to those who are unfamiliar with SLURM (Simple Linux Utility for Research Management) cluster environments and the linux command line. Therefore, here, we provide a new simple installation mechanism (conda requirements file, https://github.com/conda/conda) that requires only one installation per user. This substantially reduces the installation burden as compared to the previous pipeline, which required five program and six R package installations. To allow for this single installation containing all required programs and packages, we begin from a stable base of qiime2 2024.2n (Bolyen et al. 2019) to which we add a single R package (treetools; Smith 2019). Not only does this approach require a single installation, but the required programs and packages remain unchanged, with versions pre-set. This reduces any unforeseen version incompatibility within the pipeline. In addition, we also replaced a dependency external to qiime2 (fastqc; Andrews 2010) for read quality visualization. Instead of relying on fastqc, we now use qiime2's built in visualization commands paired with qiime's online visualization tool (view.qiime2.org).

Table 1 Improvements to the pipeline

Major improvement	Explanation
1. Backbone tree updated	Four new genera were added to the tree to reflect taxonomic discoveries.
2. Condensed ReadMe	The ReadMe has been condensed to concisely, but clearly, outline steps in the pipeline.
3. Simplified conda installation	A conda requirements file is installed once and includes all required programs and R packages in specified versions.
4. Cross computing cluster compatibility	Several changes have been made to the scripts to be compatible with the three SLURM clusters environments tested.
5. Public repository tracking changes	Our Github repository has all implemented changes tracked; we will continue to update any changes here.
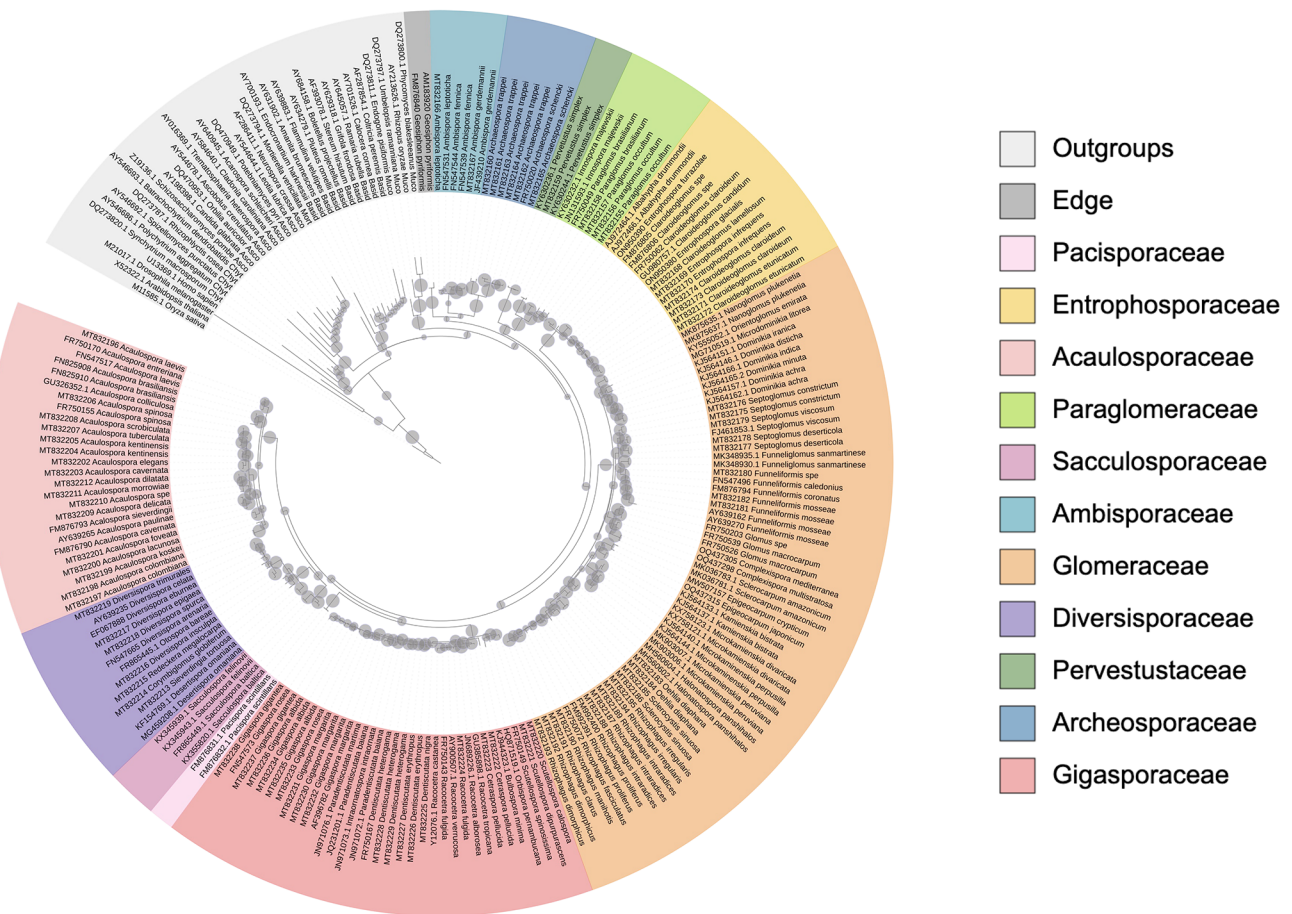
Fig. 1 Updated AMF LSU backbone tree

Put simply, this approach allows users to install one 'meta package' or folder (the conda environment) containing all packages in the exact versions required for the pipeline to run; when running the pipeline, the user simply loads this environment and can be confident that all packages and commands are exactly as intended. This new single installation will not only make it easier for all users, but especially allow for less experienced users with low cluster support to access and benefit from the pipeline.

Ground truthing ease of implementation across three computing clusters

Previous development and testing of the pipeline was conducted on only one computing cluster at the University of Kansas, raising concerns of transferability of the pipeline across different SLURM environments. Importantly, small differences in cluster setup can lead to script failure, presenting an unnecessary difficulty for those implementing the pipeline for their own research. Therefore, this current

pipeline was tested and adapted to work consistently across three different university computing clusters. Specifically, we have tested the pipeline using ten samples of data from tallgrass prairie soil samples on the (1) University of Kansas, (2) University of Colorado Boulder (Alpine), and (3) ETH Zurich (Euler) computing clusters. The ETH Zurich cluster does not support or install conda environments for users at this time, making it the ideal test for the newly integrated conda environment. The successful implementation of the pipeline on a cluster with no conda support gives us confidence that it should function smoothly at other clusters. We retain the ability to choose ASV or OTU outputs – via two versions of the pipeline – and ensure both pipelines have consistent naming and updates. Another change we implemented was explicitly directing temporary files, as this was causing issues on certain clusters (due to cluster level default differences). All temporary files should now consistently write to the main project folder, and be removed upon completion, independent of cluster configuration. Together, this testing across two continents and three

university computing clusters increases transferability and use of the pipeline.

Conclusions

We present an updated user-friendly database and pipeline for environmental placement of arbuscular mycorrhizal fungi using the LSU rDNA gene region. Our major improvements include (1) providing an expanded backbone reference tree to reflect recently described taxa (2) streamlining the requirements for users to implement the pipeline on a cluster relying on the most common workload manager (SLURM) and (3) minimizing likelihood of cluster specific issues by testing the pipeline with the same dataset across three computing clusters. These improvements may further increase the utility of the LSU region for identification of AMF and AMF families of known and previously undescribed AMF.

Author contributions CD, RR, and JB conceived of the study, CD, RR, SS, JB performed research, CD and RR analyzed data, CD lead paper writing with substantial input from RR and revisions by SS and JB.

Data availability All sequencing data has been uploaded to NCBI Bio-Project PRJNA1106240. All code for the AMF LSU pipeline is available at Github repository https://github.com/c383d893/AMF-LSU-Database-and-Pipeline2.

Declarations

Competing interests The authors declare no competing interests.

References

Anaconda (2024) Anaconda Software Distribution

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom

Bidartondo M, Bruns T, Blackwell M, Edwards I, Taylor A, Horton T, Zhang N (2008) Preserving accuracy in GenBank. Science 319:1616

Błaszkowski J, Jobim K, Niezgoda P, Meller E, Zubek S, Magurno F, Casieri L, Bierza W, Błaszkowski T, Crossay T (2021a) New Glomeromycotan Taxa, Dominikia Glomerocarpica sp. nov. and Epigeocarpum Crypticum gen. nov. et sp. nov. From Brazil, and Silvaspora gen. Nov. From New Caledonia. Front Microbiol 12:655910

Błaszkowski J, Niezgoda P, Meller E, Milczarski P, Zubek S, Malicka M, Uszok S, Casieri L, Goto BT, Magurno F (2021b) New Taxa in Glomeromycota: Polonosporaceae fam. nov., Polonospora gen. nov., and P. Polonica comb. Nov. Mycological Progress 20:941–951

Błaszkowski J, Sánchez-García M, Niezgoda P, Zubek S, Fernández F, Symanczik S, Malinowski R, Cabello M, Goto BT, Malicka M (2022) A new order, Entrophosporales, and three new Entrophospora species in Glomeromycota. Front Microbiol 13:962856

Błaszkowski J, Yamato M, Niezgoda P, Zubek S, Milczarski P, Malinowski R, Meller E, Malicka M, Goto BT, Uszok S (2023) A new genus, Complexispora, with two new species, C. Multistratosa and C. Mediterranea, and Epigeocarpum japonicum sp. nov. Mycological Progress 22:34

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857

Bunyard BA, Nicholson MS, Royse DJ (1994) A systematic assessment of Morchella using RFLP analysis of the 28S ribosomal RNA gene. Mycologia 86:762–772

da Silva GA, Corazon-Guivin MA, de Assis DMA, Oehl F (2023) Blaszkowskia, a new genus in Glomeraceae. Mycological Progress 22:74

Delavaux CS, Sturmer SL, Wagner MR, Schütte U, Morton JB, Bever JD (2021) Utility of large subunit for environmental sequencing of arbuscular mycorrhizal fungi: a new reference database and pipeline. New Phytol 229:3048–3052

Delavaux CS, Ramos RJ, Sturmer SL, Bever JD (2022) Environmental identification of arbuscular mycorrhizal fungi using the LSU rDNA gene region: an expanded database and improved pipeline. Mycorrhiza 32:145–153

Hart MM, Aleklett K, Chagnon PL, Egan C, Ghignone S, Helgason T, Lekberg Y, Öpik M, Pickles BJ, Waller L (2015) Navigating the labyrinth: a guide to sequence-based, community ecology of arbuscular mycorrhizal fungi. New Phytol 207:235–247

House GL, Ekanayake S, Ruan Y, Schütte UM, Kaonongbua W, Fox G, Ye Y, Bever JD (2016) Phylogenetically structured differences in rRNA gene sequence variation among species of arbuscular mycorrhizal fungi and their implications for sequence clustering. Appl Environ Microbiology:AEM. 00816–00816

Krüger M, Stockinger H, Krüger C, Schüßler A (2009) DNA-based species level detection of Glomeromycota: one PCR primer set for all arbuscular mycorrhizal fungi. New Phytol 183:212–223

Lekberg Y, Vasar M, Bullington LS, Sepp SK, Antunes PM, Bunn R, Larkin BG, Öpik M (2018) More bang for the buck? Can

arbuscular mycorrhizal fungal communities be characterized adequately alongside other fungi using general fungal primers? New Phytol 220:971–976

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

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci 109:6241–6246

Smith MR (2019) TreeTools: create, modify and analyse phylogenetic trees. Compr R Archive Netw 10

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Stockinger H, Krüger M, Schüßler A (2010) DNA barcoding of arbuscular mycorrhizal fungi. New Phytol 187:461–474

Tedersoo L, Bahram M, Zinger L, Nilsson RH, Kennedy PG, Yang T, Anslan S, Mikryukov V (2022) Best practices in metabarcoding of fungi: from experimental design to results. Mol Ecol 31:2769–2795

Trouvelot S, van Tuinen D, Hijri M, Gianinazzi-Pearson V (1999) Visualization of ribosomal DNA loci in spore interphasic nuclei of glomalean fungi by fluorescence in situ hybridization. Mycorrhiza 8:203–206

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.