Global Ecology and Biogeography

WILEY

Global Ecology and Biogeography | A Journal of Macroecology

DATA ARTICLE OPEN ACCESS

AusAMF: The Database of Arbuscular Mycorrhizal Fungal Communities in Australia

Adam Frew[1,2,3] | Jeff R. Powell[1] | Meike K. Heuck[1] | Felipe E. Albornoz[4,5,6] | Christina Birnbaum[2,7,8] | John D. W. Dearnaley[2,7] | Eleonora Egidi[1] | Luke Finn[9] | Jarrod Kath[7,8] | Kadri Koorem[10] | Jane Oja[10] | Maarja Öpik[10] | Tanel Vahter[10] | Martti Vasar[10] | Stephanie Watts-Williams[11] | Yuxiong Zheng[1,12,13] | Carlos A. Aguilar-Trigueros[1,3]

[1]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales, Australia | [2]Centre for Crop Health, University of Southern Queensland, Toowoomba, Queensland, Australia | [3]Department of Biological and Environmental Sciences, University of Jyväskylä, Jyväskylä, Finland | [4]Commonwealth Scientific and Industrial Research Organisation, Environment, Waterford, Western Australia, Australia | [5]School of Environmental and Conservation Sciences, Murdoch University, Murdoch, Western Australia, Australia | [6]School of Biological Sciences, The University of Western Australia, Perth, Western Australia, Australia | [7]School of Agriculture and Environmental Science, University of Southern Queensland, Toowoomba, Queensland, Australia | [8]Centre for Sustainable Agricultural Systems, University of Southern Queensland, Toowoomba, Queensland, Australia | [9]Terrestrial Ecosystem Research Network, The University of Adelaide, Adelaide, South Australia, Australia | [10]Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia | [11]School of Agriculture, Food and Wine, the Waite Research Institute, The University of Adelaide, Adelaide, South Australia, Australia | [12]Key Laboratory of low-Carbon Green Agriculture in Northwestern China, Ministry of Agriculture and Rural Affairs of China, College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi, China | [13]State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling, Shaanxi, China

Correspondence: Adam Frew (a.frew@westernsydney.edu.au)

Received: 14 November 2024 | **Revised:** 4 May 2025 | **Accepted:** 3 July 2025

Handling Editor: Margaret Morrow Mayfield

Funding: The majority of this work was supported by an Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA; DE220100479) awarded to A.F. and which also supported M.K.H. through a postgraduate research scholarship. J.R.P. was supported by an ARC Future Fellowship (FT190100590), S.J.W.-W. was supported by an ARC DECRA (DE210100908), as was E.E. (DE210101822). C.A.A-T was supported by an Academy Research Fellowship (21000058691) from the Research Council of Finland. T.V, M.Ö., and K.K. were supported by the Estonian Research Council (grants PRG1789, PRG1836) and the Centre of Excellence AgroCropFuture. J.O. was supported by the Estonian Research Council (PSG784).

Keywords: arbuscular mycorrhizal fungi | Australia | DNA metabarcoding | high-throughput sequencing | plant symbionts | soil fungi | symbiosis

ABSTRACT

Motivation: Arbuscular mycorrhizal (AM) fungi are central to plant nutrient acquisition, soil carbon dynamics, and ecosystem resilience. Yet, their biogeography remains incompletely characterised, particularly across underrepresented regions. Australia, with its characteristic ecological conditions, continental scale, and long-standing evolutionary trajectories, has been notably undersampled. This gap hinders our ability to make comprehensive inferences about AM fungal diversity, community composition, and ecological roles at global scales. The AusAMF database was created to address this deficiency by compiling high-throughput AM fungal community data across mainland Australia and Tasmania. The initial release comprises data from 610 georeferenced sites sampled between 2011 and 2023, covering all major climate zones and accompanied by standardised soil storage, DNA extraction, and sequencing procedures. Developed through a nationally coordinated effort, AusAMF offers a rare level of methodological consistency, enabling robust spatial and temporal comparisons while minimising post-sampling technical biases. Its design as a purpose-built, extensible platform ensures continued expansion using harmonised protocols—something not achieved through compiled datasets assembled retrospectively from disparate studies. Each sample is linked to associated environmental variables, allowing users to explore ecological drivers of AM fungal distributions, assess patterns of biodiversity, and

support applications spanning from fundamental ecology to conservation planning. As such, AusAMF advances both regional and global efforts to characterise the diversity and ecological significance of these foundational plant symbionts.

Main Types of Variables Contained: Georeferenced occurrence and abundance of high-throughput amplicon sequences of arbuscular mycorrhizal fungi.

Spatial Location and Grain: Australia. Decimal degrees between 0.0001 and 0.1 resolution.

Time Period and Grain: 2011–2023. Month and year of sampling.

Major Taxa and Level of Measurement: Arbuscular mycorrhizal fungi identified to family, genus, and virtual taxon (VT). Geographic occurrence and amplicon sequence abundance.

Software Format: Interact with processed data via online application (https://www.ausamf.com). Dataset available as .csv files and raw sequencing data as .fastq files.

1 | Introduction

The majority of terrestrial plants form symbiotic relationships with arbuscular mycorrhizal (AM) fungi (Smith and Read 2008) which colonise plant roots and extend extensive mycelial networks into the soil. As obligate symbionts, these fungi depend entirely on carbon supplied by the host, which can constitute as much as 17% of the total carbon fixed by some plants (Harris et al. 1985). The fungi facilitate plant access to essential nutrients, particularly phosphorus (P), where the supply via AM fungi can represent up to 80% of the host's P nutrition (Marschner and Dell 1994). Yet the role of AM fungi extends far beyond nutrient uptake. They influence plant community assembly, mediate resistance to abiotic and biotic stresses, and drive carbon and nutrient cycling across terrestrial ecosystems (Read and Perez-Moreno 2003; Delavaux et al. 2017; Neuenkamp et al. 2018; Frew et al. 2022). Given their importance, adequate sampling across key global regions is essential to revealing how their roles vary across ecosystems (Davison et al. 2015). Such knowledge is necessary not only for elucidating ecological interactions and co-evolutionary dynamics, but also for modelling global processes such as carbon cycling and ecosystem resilience under environmental change.

Despite the importance of AM fungi, our current understanding of their biogeography is geographically skewed. Global datasets are dominated by studies from Europe, North America, and China. The GlobalAMFungi database (version 1.0; Větrovský et al. 2023), collating global AM fungal composition data from 100 studies, highlights pronounced undersampling in the Southern Hemisphere. Australia, despite encompassing more than 5% of the world's land area, is currently represented by 32 soil samples in the GlobalAMFungi database (0.8% of the total). This limited sampling provides little basis for understanding AM fungal diversity across the continent's environmental gradients and varied ecosystems.

Improving representation from Australia is critical to resolving this geographic imbalance and capturing fungal diversity from regions underrepresented in global datasets. Its ancient soils, nutrient-impoverished landscapes, and highly endemic flora offer an opportunity to investigate AM fungal communities under environmental conditions largely absent from Northern Hemisphere studies (Orians and Milewski 2007; Lambers et al. 2011; Teste et al. 2020). Moreover, the continent's Gondwanan legacy links its ecosystems to those of Africa and South America, suggesting that insights gleaned from Australia may carry broader biogeographic implications (Flores-Moreno et al. 2023). Many Australian plant species form multiple types of mycorrhizal associations—or none at all—suggesting potential divergences in host-symbiont compatibility, selectivity, and functional outcomes relative to other continents. To overlook these patterns is to risk assuming universality where there may be deep contextual variation (Teste et al. 2020).

While molecular techniques have revolutionised the study of fungal communities, AM fungi require a more nuanced methodological approach than other fungal groups (Větrovský et al. 2023). For the majority of fungi, research on their diversity can be adequately achieved using the internal transcribed spacer (ITS) region of the rRNA gene cluster, recognised as the best suited marker for investigating community composition and diversity in most fungal lineages through amplicon sequencing (DNA metabarcoding) (Schoch et al. 2012). However, AM fungi possess highly variable ITS copies, leading to probable overestimations of species richness and strong biases towards certain lineages in studies that solely rely on ITS primers (Kohout et al. 2014; Öpik et al. 2014; Lekberg et al. 2018).

Consequently, the 18S small subunit (SSU) and the 28S large subunit (LSU) regions of the rRNA gene cluster are considered more appropriate markers for describing AM fungal diversity (Stockinger et al. 2010; Kohout et al. 2014). Although the debate regarding the choice of barcode markers is ongoing (Kohout et al. 2014; Lekberg et al. 2018; Delavaux et al. 2021; Tedersoo et al. 2022), most contemporary studies employ sequencing of the 18S small subunit (SSU) rRNA gene region, which is less variable than the ITS. The SSU offers conservative yet robust estimates of AM fungal diversity and phylogenetic structure (Öpik et al. 2014; Thiéry et al. 2016). Additionally, the SSU is the primary marker utilised by MaarjAM (Öpik et al. 2010), a curated database of AM fungal SSU sequences, which links sequences to phylogenies and assigns them to reference 'virtual taxon/taxa' (VT), thereby enhancing comparability across different studies. As the SSU region is slow-evolving, potentially limiting the ability to resolve AM fungal species, there is growing support for more widespread adoption of the LSU region for environmental sequencing of AM fungi (Delavaux et al. 2021). This region has historically been utilised less, ostensibly due to bioinformatical challenges, but may become more useful as reference databases and user-friendly pipelines continue to develop (Delavaux et al. 2024).

Here we introduce the AusAMF database, the first continental-scale compilation dedicated to AM fungal community data across Australia and Tasmania (Figure 1). This initial release includes samples from 610 locations collected between 2011 and

2023, spanning all major climate zones (Figure 2), making the dataset well-suited for inferring biogeographic patterns of AM fungi. Unlike contributions to global repositories such as the GlobalAMFungi database, which aggregate data from studies with heterogeneous methodologies, AusAMF provides a nationally coordinated dataset with a high degree of methodological consistency in downstream processing. All soil samples were air-dried and stored with silica to maintain dryness prior to DNA extraction, which was carried out using the same protocol across samples. Sequencing of the SSU region was performed using the WANDA (Dumbrell et al. 2011) and AML2 (Lee et al. 2008) primer set on the Illumina NextSeq platform (2×300 bp) at a single facility. This standardisation in post-sampling procedures is essential for robust spatial comparisons and longitudinal analyses, minimising technical biases that might otherwise obscure ecological patterns. The raw sequencing data, in FASTQ format, are publicly available via the Sequence Read Archive (accession PRJNA1154677).

AusAMF is designed as an ongoing, expandable resource—future updates will maintain these methodological standards while incorporating new samples across time and space. By addressing a longstanding spatial imbalance in global AM fungal sampling through this harmonised framework, AusAMF enables researchers to test new hypotheses about how environmental, biogeographic, and evolutionary factors shape fungal communities within Australia and in comparative global contexts.

In addition to the raw sequence data, the database includes data processed using gDAT (Vasar et al. 2021), a bioinformatic pipeline specifically optimised for analysis of SSU amplicons from AM fungi. Taxonomy is assigned to VT using the MaarjAM database (Öpik et al. 2010). For each sample, the spatial coordinates of sampling locations were used to download climate and soil variables from public repositories (WorldClim, Soil and Landscape Grid of Australia). These environmental data were combined with the AM fungal community data to allow users easy access to information that can be relevant for further analyses.

AusAMF will continue to grow as new AM fungal community data become available. Contributions from the wider research

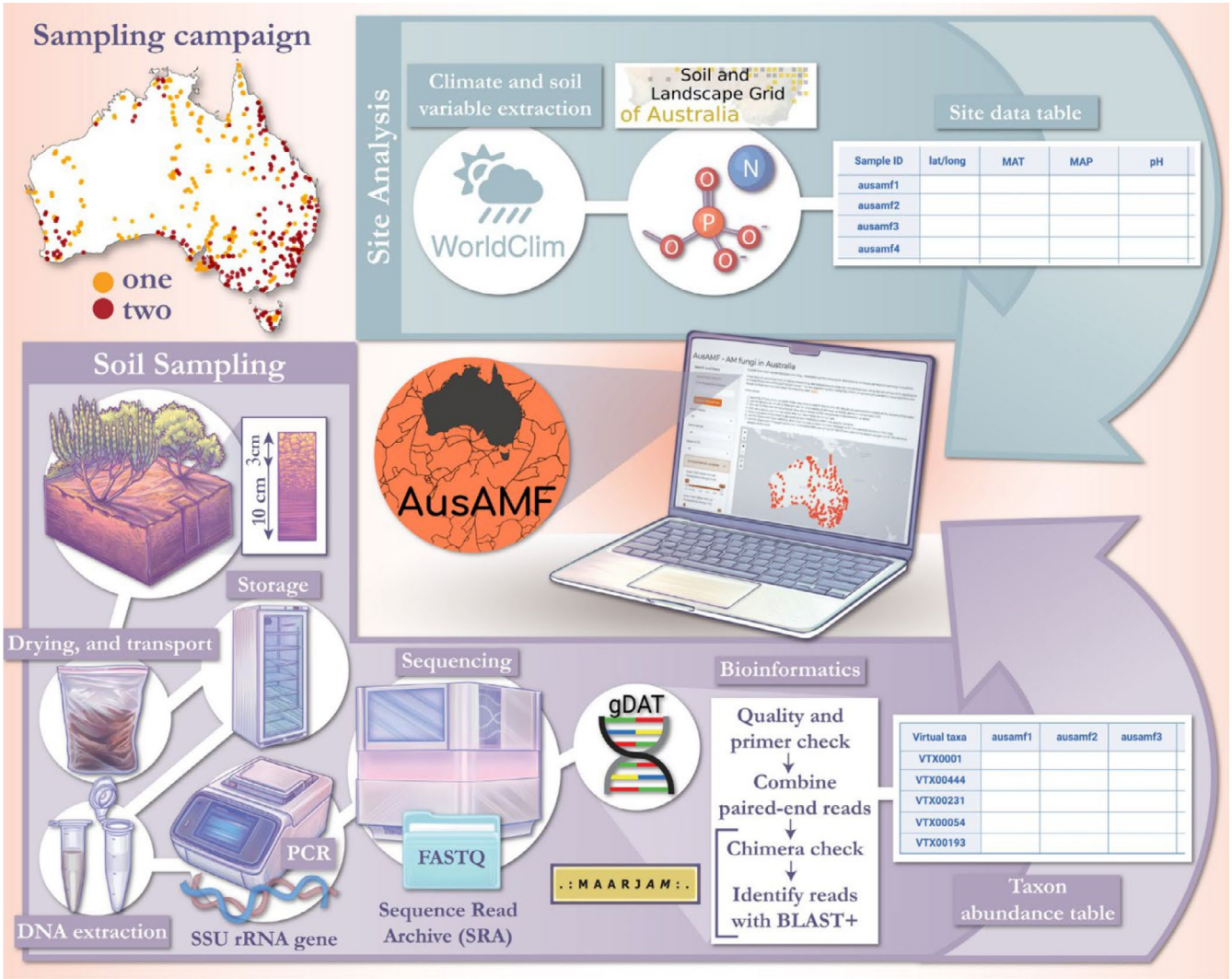

FIGURE 1 | Workflow in generating data for AusAMF. Climate and soil variables for each sample location are sourced from WorldClim and the Soil and Landscape Grid of Australia. These data are combined with AM fungal community sequencing data for each location and made available through the online AusAMF database (www.ausamf.com).

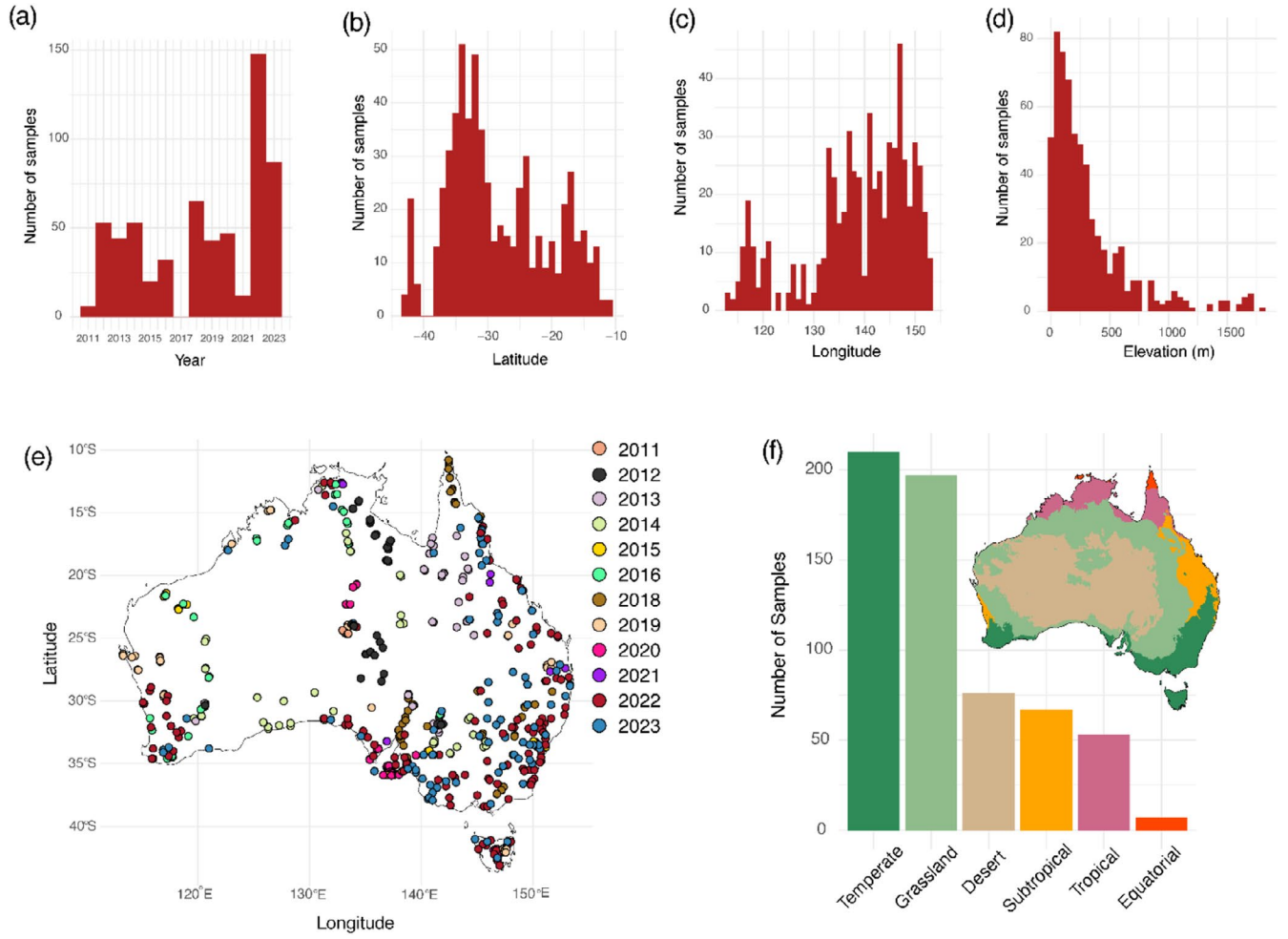

FIGURE 2 | The distribution of samples in AusAMF by (a) year of sampling (b) latitude (c) longitude (d) elevation. (e) Map of sample locations coloured by year of sampling and (f) the number of samples from each of the major climate zones.

community are welcomed, provided samples, extracts, or datasets are compatible with the existing methodological framework. In this way, AusAMF will serve as an expanding resource for spatial and temporal data on AM fungal communities, supporting integrative ecological research and conservation planning.

2 | Methods

2.1 | Soil Sampling

The Australian arbuscular mycorrhizal fungal database (AusAMF) comprises high-throughput SSU gene region sequence data from samples of soil collected across the Australian mainland and the island of Tasmania. Soil sampling for this initial dataset release (version 1.0) was conducted through two major sampling campaigns spanning from 2011 to 2023.

The first sampling campaign utilised sites from the Terrestrial Ecosystem Research Network (TERN) between 2011 and 2021, following the methods comprehensively described in White et al. (2012). In brief, loose debris was moved from the soil surface and a soil sample was collected from the top 3 cm of the soil profile at a sampling point across 378 locations (100 × 100 m plots) in Australia included in the first release of this dataset (Figure 1). During sampling, gloves were worn, and instruments were cleaned with ethanol, methylated spirits, or 20% commercial bleach between samples to prevent contamination across samples and sites. Soils were air-dried and stored with silica gel in sealed bags or screw-top sample containers to maintain dryness. These samples were stored at a temperature-controlled facility (18°C) until they were transported to Western Sydney University (Richmond, NSW, Australia) in 2023, where they were stored with silica between 4°C–5°C until DNA extraction.

The second sampling campaign, conducted over 12 months from August 2022 to September 2023, involved soil collection from 230 sites, spanning both agricultural and non-agricultural environments across Australia. At each site, six soil sub-samples were collected from within the top 10 cm of the soil within a 25 × 25 m plot and composited. This depth was selected to account for potential soil turnover and disturbance, particularly in agricultural environments where many of the samples originated. These subsamples were collected at equal distances along a transect, or if not possible, subsamples were collected equally spaced in a grid from the 25 m^2 plot. Samples were collected wearing gloves; litter and debris were removed from the surface prior to taking the soil. Any equipment and instruments used were cleaned with ethanol, methylated spirits, or at least 20% commercial bleach, prior to sampling. The soil samples

were air-dried at room temperature (avoiding direct sunlight and temperatures above 39°C) and sealed in plastic bags with silica gel before being transported to Western Sydney University (Richmond, NSW, Australia), where they were stored at 4°C until DNA extraction.

2.2 | Environmental Variables

Site locations as latitude and longitude were recorded at sampling. Those samples collected from sampling campaign one (TERN) provide locations of sampling points at a resolution of 0.0001 decimal degrees (~10 m); soil collections from the second campaign were provided by third parties comprising various researchers, land managers, and volunteers. The longitude and latitude for these locations are provided at a lower resolution of 0.1 decimal degrees (~10 km). Environmental variables included in the dataset are mean annual temperature (MAT), mean annual precipitation (MAP), elevation, soil total nitrogen, soil total phosphorus, soil pH ($CaCl_2$), soil organic carbon (SOC), and soil available phosphorus.

The MAT, MAP, and elevation data for the locations were retrieved from the WorldClim Bioclimatic variables (Fick and Hijmans 2017), which provide environmental data based on the average for the years 1970–2000. The soil variables were retrieved from the Soil and Landscape Grid of Australia, an open-access database using digital soil mapping and extensive soil sampling across the continent (Terrestrial Ecosystem Research Network 2024). Details of the sampling, analyses, and mapping methodologies are available for total nitrogen (Rossel et al. 2014), total phosphorus (Malone and Searle 2024), pH (Malone and Searle 2021), soil organic carbon (Wadoux et al. 2022), and available (Colwell) phosphorus (Zund 2022).

2.3 | DNA Extraction and Sequencing

The DNA from each soil sample was extracted from 250 mg of soil using DNeasy Powersoil Pro Kits (Qiagen, GmBH, Hilden, Germany) according to the manufacturer's instructions, apart from the final step where DNA was eluted using UltraPure Distilled Water (Invitrogen, MA, USA) to minimise interference of buffers with downstream sequencing steps. DNA underwent amplification by polymerase chain reaction (PCR) of the small-subunit (SSU) ribosomal RNA gene using the WANDA (5′-CAGCCGCGGTAATTCCAGCT-3′; Dumbrell et al. 2011) and AML2 (5′-GAACCCAAACACTTTGGTTTCC-3′; Lee et al. 2008) primer set with Illumina overhang sequences. Sequencing was performed by the Ramaciotti Centre for Genomics (UNSW Sydney, Australia). For this, thermal cycling conditions for PCR were as follows: 95°C for 2 min; 27 cycles of 95°C for 60 s, 54°C for 60 s, and 72°C for 60s; with a final extension of 72°C for 10 min. DNA quality control gels were used to confirm successful amplification. All samples were purified using the Agencourt AMPure XP Beads (Beckman Coulter, CA, USA). The PCR products were then amplified in an indexing PCR reaction with Nextera-compatible unique dual indexes (Integrated DNA Technologies, Coralville, IA, USA). The thermal cycling conditions were: 95°C for 3 min; 8 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30s; with a final extension of 72°C for 5 min. Quality control gels were again used to confirm successful amplification, and all samples were normalised using SequalPrep Normalisation Plate Kit A1051001 (Thermo Fisher Scientific, MA, USA). All samples were pooled with an equal volume and cleaned using the Agencourt AMPure XP Beads (Beckman Coulter, CA, USA). The cleaned sample pool was quantified by Qubit quantification assay (Thermo Fisher Scientific, MA, USA) before being sequenced on an Illumina NextSeq 1000 platform with a 2×300 bp paired-read approach using the Illumina P1 600 XL kit.

2.4 | Bioinformatics

Bioinformatic data analysis and processing was conducted using the graphical downstream analysis tool (gDAT) for analysing rDNA sequences (Vasar et al. 2021). The default settings of the gDAT pipeline are optimised for the analysis of AM fungal SSU sequences. We followed the defaults as outlined in Vasar et al. (2021). The raw paired-end reads (2×88,022,138) were retained if they carried the correct primer sequences (WANDA and AML2; allowing one mismatch for each). Quality filtering was done by checking the average base quality against a threshold (at least 30) combined with a sliding window average quality check to trim low-quality regions, leaving 2×71,464,445 cleaned reads. Paired-end reads were combined using FLASH (Magoč and Salzberg 2011) with default minimum 10-bp overlap and 75% overlap identity threshold, resulting in 69,447,662 (97.18%) combined pairs. Although gDAT provides the option, we did not pre-cluster reads into operational taxonomic units (OTUs). Chimera checking was conducted with VSEARCH (Rognes et al. 2016) in denovo and reference mode which checked each sequence individually against the reference database, in this case, MaarjAM database (Öpik et al. 2010). This identified 163,094 putative chimeras (0.2%) and 69,258,958 non-chimeras (99.7%). The AM fungal sequences, and their taxonomies, were identified using BLAST+ (Camacho et al. 2009) where each individual read was identified in comparison to the MaarjAM database (Öpik et al. 2010) as a reference using the nucleotide local pairwise alignment BLASTN algorithm. This identified sequences to phylogenetically defined taxonomic units called virtual taxa (VT). A taxon abundance table was then generated at 97% identity matching and 95% read alignment.

2.5 | Data Exploration and Visualisation

All data exploration and analyses were conducted using R version 4.3.3 (R Core Team 2024).

The MAP, MAT, elevation, soil pH ($CaCl_2$), soil organic carbon, total nitrogen, total phosphorus, and available phosphorus were extracted for each sample location using the 'raster' (Hijmans 2023) and 'sp' (Bivand and Gomez-Rubio 2013) packages in R. The raster layers used for these extractions have been compiled into a spatially aligned environmental datacube and are available for users to access via the Figshare repository (Frew et al. 2025). Frequency histograms, scatterplots, mean and standard error plots were generated using 'ggplot2' (Wickham 2016). Maps of Australia were produced using the 'rnaturalearth' package (Massicotte and South 2023).

Rarefaction curves to compare sequencing depth to number of VT were made using *rarecurve* from 'vegan' (Oksanen et al. 2015). Sample completeness was quantified using the 'iNEXT' package (Hsieh et al. 2016), which estimates sample coverage as the proportion of the total community captured by observed sequences. For the purposes of describing patterns of diversity and composition (Figures 3 and 4), we calculated diversity estimates using Hill numbers (Hill numbers; Chao et al. 2014) standardised at 95% sample coverage to account for variation in sequencing depth and enable robust comparisons across samples with minimal extrapolation. To visualise patterns in community composition, proportional abundance plots were generated after applying variance-stabilising transformation using 'DESeq2' (Love et al. 2014), which adjusts for differences in library size while preserving relative abundance structure. These transformed counts were then visualised using the *transform_sample_counts* and *psmelt* functions from 'phyloseq' (McMurdie and Holmes 2013) together with 'ggplot2' (Wickham 2016).

The online interactive application for AusAMF was created using the 'shiny' package in R (Chang et al. 2024), together with packages 'shinyBS' (Bailey 2022), 'bslib' (Sievert et al. 2024), 'leaflet' (Cheng et al. 2024), 'DT' (Xie et al. 2024),

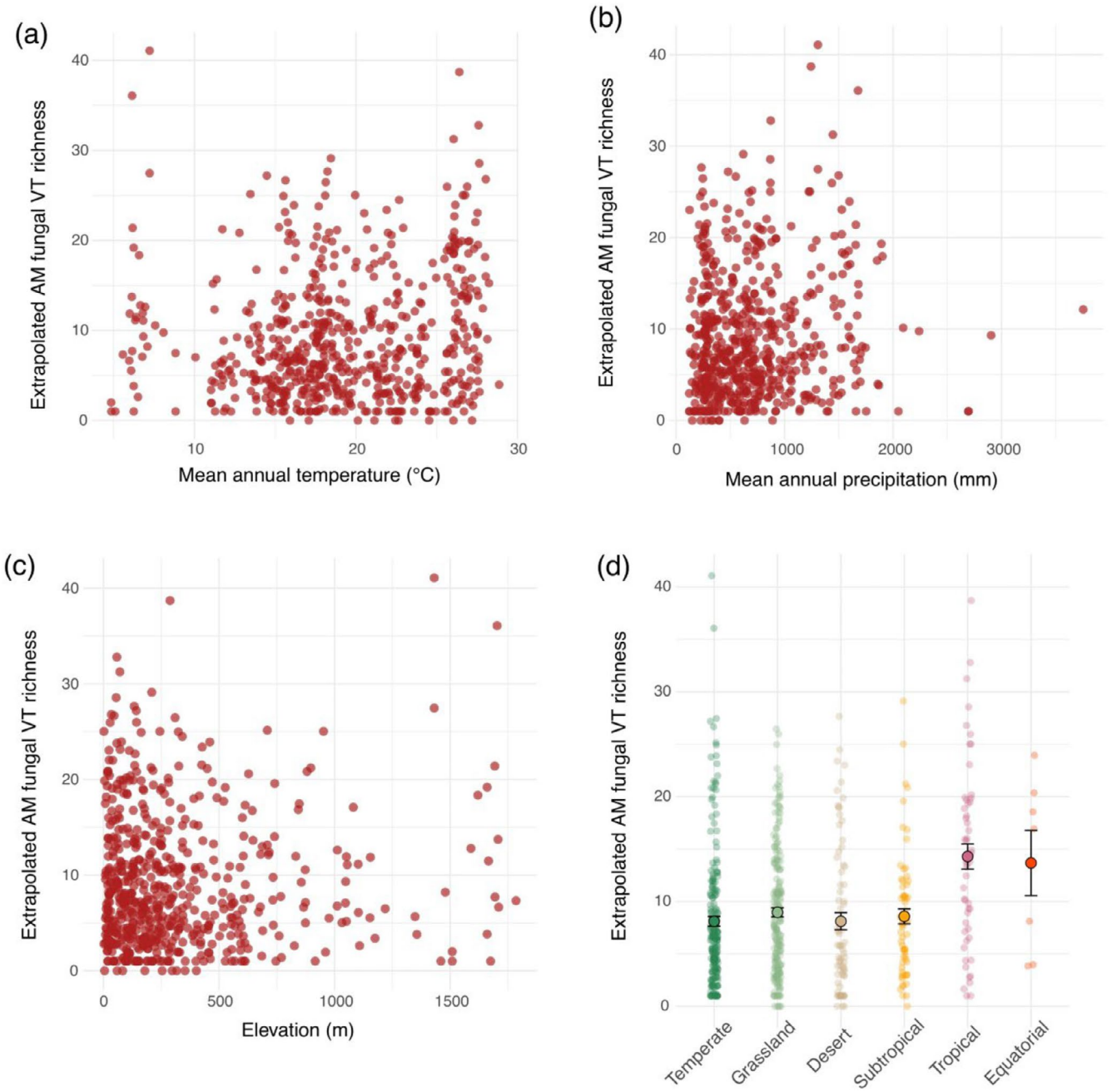

FIGURE 3 | Scatterplots showing relationships between arbuscular mycorrhizal (AM) fungal virtual taxon (VT) richness and (a) mean annual temperature (b) mean annual precipitation, (c) elevation, and (d) major climate zone based on the modified Köppen–Geiger classification. Richness data are extrapolated VT richness (Hill number q = 0) standardised at 98% sample coverage to account for variation in sequencing depth. Solid points and error bars in (d) represent mean ± standard error which are overlaid on top of the raw data points.

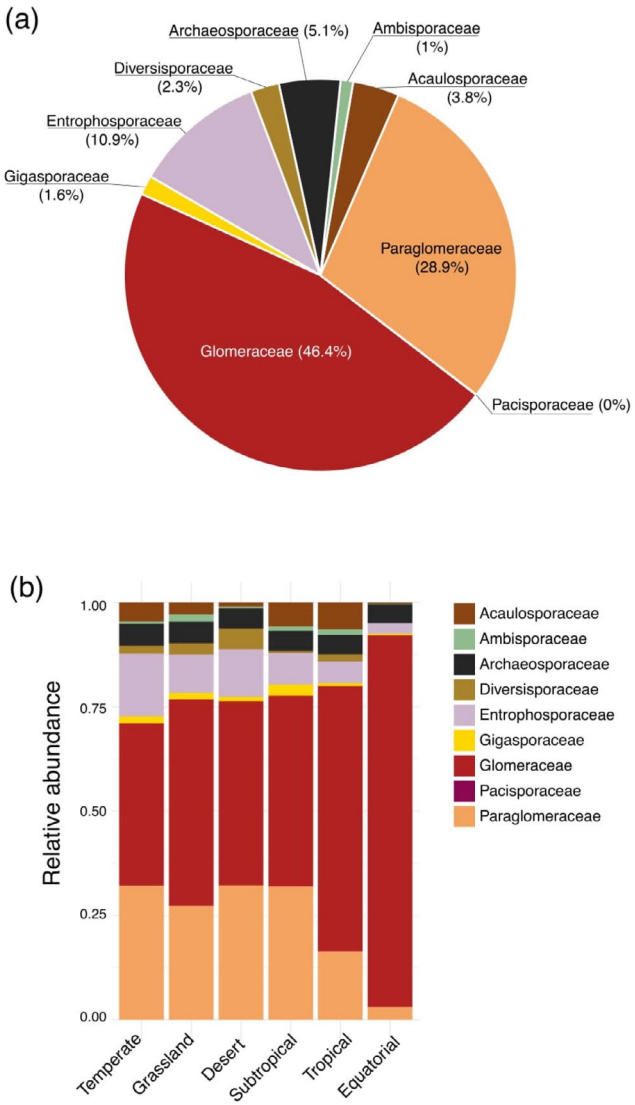

FIGURE 4 | Relative abundances of arbuscular mycorrhizal (AM) fungal families within (a) AusAMF database and (b) the relative abundance of AM fungal families in the major climate zones based on modified Köppen–Geiger classification. Abundances shown here are based on size-factor normalised counts adjusted for sequencing depth.

'dplyr' (Wickham et al. 2023), 'ape' (Paradis and Schliep 2019), 'Biostrings' (Pagès et al. 2024), 'ggtree' (Yu et al. 2017), 'readr' (Wickham et al. 2024) and 'ggplot2' (Wickham 2016).

3 | Data Description

3.1 | Sample Coverage

AusAMF release 1.0 includes high-throughput amplicon sequencing data from 610 sites in Australia sampled between 2011 and 2023 (Figure 2a,e). Geographic coverage was slightly superior in the higher absolute latitudes (i.e., southern latitudes) and higher longitudes (Figure 2b,c). The majority of sample sites sit <500 m in elevation, which reflects most of the Australian continent (87% sits below 500 m; McKenzie et al. 2004). Site locations were classified to a climate zone based on a modified Köppen-Geiger classification system which incorporates information on the climatic limits of native vegetation, as the best expression of climate in an area (Stern et al. 2000). Temperate and grassland climate zones had the most sample coverage in the dataset (Figure 2f).

3.2 | Amplicon Sequencing and Database

After quality filtering, primer and chimera checking, a total of 69,258,958 amplicon reads were obtained for further analysis. The mean and median number of sequences per sample was 113,539 and 73,975, respectively (Figure 5). The individual sequences were identified to VT referencing the MaarjAM database (Öpik et al. 2010) which identified 2,336,123 reads across the samples from 610 sites.

In our dataset we identified 200 AM fungal VT across all samples, comprising nine families. Rarefaction curves comparing the number of sequences to the number of VT exhibited plateauing in most samples, suggesting an adequate sequencing depth in most instances to capture AM fungal diversity. Assessment of sampling completeness showed that sample coverage was uniformly high across the dataset, with a mean of 99.8%, a median of 100%, and a minimum of 90.5%. These values indicate excellent completeness in most samples.

The mean extrapolated (Hill $q=0$) VT richness (Figure 3) was highest in tropical (14.3 ± 1.2; mean $\pm$ SE) and equatorial samples (13.7 ± 3.11; mean $\pm$ SE), while the lowest mean richness was in temperate samples (8.11 ± 0.46; mean $\pm$ SE). In terms of compositional relative abundance (Figure 4), the majority of sequences were from the family Glomeraceae (46.4%), while Pacisporaceae represented the lowest relative abundance of sequences ($<0.1\%$). The data within AusAMF are not rarefied or extrapolated; however, researchers should carefully consider if and how the uneven sequencing depth among samples might affect their research question(s) and how they might deal with this when conducting their own analyses.

3.3 | Database Access

AusAMF can be accessed via the graphical user interface of the online application available at www.ausamf.com. From this platform, users can download a CSV file containing processed sequencing data for each sample along with associated environmental variables (Table 1). The application also provides options for visualising and filtering samples and AM fungal taxa, which can then be exported as a CSV file. For instance, users may visualise and download data for samples containing specific taxa (at the family, genus, or VT level) or filter samples based on environmental variables such as MAT, MAP, or soil pH. Summary data are generated within the user interface when samples are selected, either by clicking on individual samples or by selecting samples on the map using the polygon tool. A phylogenetic tree is also generated for the AM fungal VT present in the selected sample(s). Additionally, the platform offers a search function that allows users to input a query sequence, returning a table of representative VT sequences from the AusAMF dataset ranked by their similarity

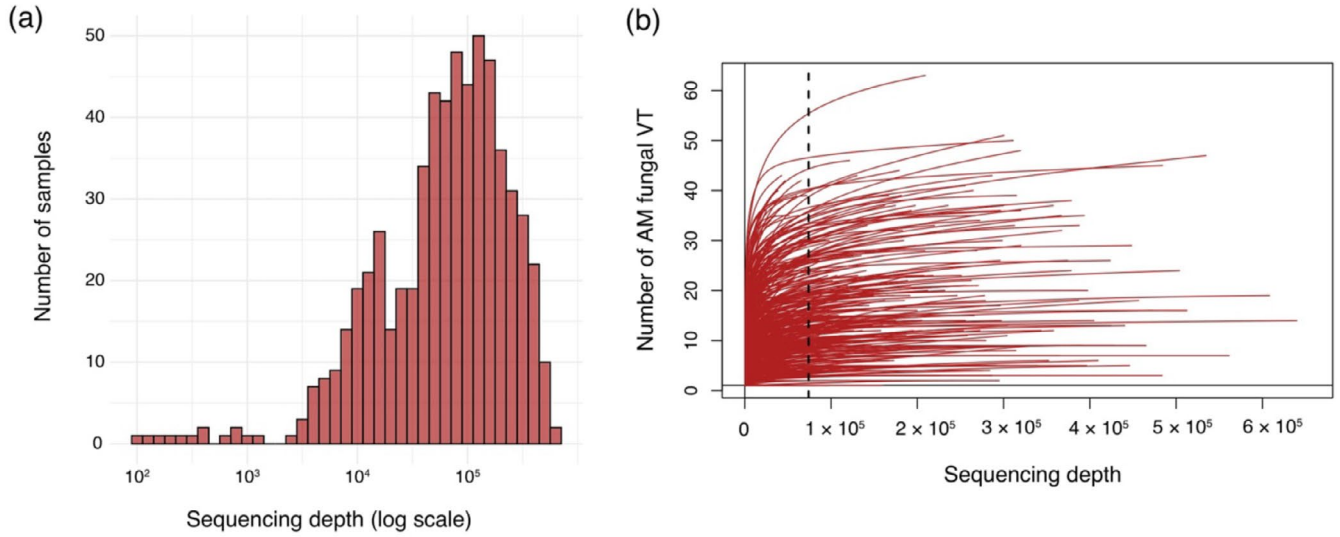

FIGURE 5 | (a) The number of samples by sequencing depth (quality filtered reads) and (b) the rarefaction curves for all samples showing accumulative number of arbuscular mycorrhizal (AM) fungal virtual taxa (VT) by sequencing depth, the median sequencing depth is shown by the dashed line.

TABLE 1 | Summary of information included in the AusAMF database.

Heading	Description
sample_number	Sample designation for the AusAMF database. These all start with ausamf followed by a numerical designation ### (e.g., ausamf015, ausamf524)
FASTQ_name	Sample name on the FASTQ files associated with the sample
latitude	The latitude of the location where the sample was taken
longitude	The longitude of the location where the sample was taken
year	The year the soil sample was collected
month	The month the soil sample was collected
koppen	The major climate zone classification, based on modified Köppen-Geiger classification system
MAT	Mean annual temperature (°C), calculated average between the years 1970–2000, for the location the sample was taken
MAP	Mean annual precipitation (mm), calculated average between the years 1970–2000, for the location the sample was taken
elevation	The elevation (m) for the location the sample was taken
total_N	Total nitrogen (%) in soil of the location the sample was taken
total_P	Total phosphorus (%) in soil of the location the sample was taken
avail_P	Available (Colwell) phosphorus (mg/kg) in soil of the location the sample was taken
SOC	Soil organic carbon (%) in soil of the location the sample was taken.
pH_cacl	Soil pH ($CaCl_2$) of the location the sample was taken
sampling_campaign	The sampling campaign the soil sample was collected from. The first version of AusAMF includes samples from two campaigns (see methods). This column will have either '1' (the first sampling campaign) or '2' (the second campaign)
soil_depth	The depth (cm) from which the soil sample was taken
cleaned_reads	The total number of sequence reads in the sample after quality filtering and chimera checking
total_abundance	The number of reads of the associated AM fungal VT in the sample
VT_ID	The virtual taxa (VT) identifier for AM fungi has the prefix 'VTX' followed by five numbers (e.g., VTX00142, VTX00105)
Family	Taxonomic assignment to family of the AM fungal VT
Genus	Taxonomic assignment to genus of the AM fungal VT

to the input sequence. The raw sequencing data are available for download from the NCBI Sequence Read Archive under accession PRJNA1154677, provided as paired-end FASTQ files for each of the 610 samples included in this first release. The raster layers used to extract climate and edaphic variables have been compiled into a spatially aligned environmental datacube (multi-layer GeoTIFF format), which is provided in the Figshare repository (Frew et al. 2025). This enables reproducibility of site-level environmental data and allows users to reload and manipulate the data using standard spatial tools (e.g., terra::rast; Hijmans et al. 2022).

The breadth and structure of AusAMF not only facilitate the detection of broad-scale biogeographical patterns of AM fungi across this under-sampled continent, but also provide a rare opportunity to develop much needed hypothesis-driven research into the ecological and evolutionary processes shaping AM fungal communities at continental and local scales. With AusAMF, Australia will be the only continent for which AM fungal communities have been characterised at this scale with a consistent approach to sample processing, sequencing, and data curation. With this spatial resolution and environmental metadata, it will be possible, for example, to test whether AM fungal distributions mirror the biogeographical patterns of the Australian flora, or diverge in response to soil properties and land use history. As the database continues to grow over time, it will also become increasingly valuable for investigating temporal shifts in AM fungal communities, whether due to climate change, urban expansion, or land-use intensification. Moreover, given the substantial carbon investment made by host plants into AM fungi (Hawkins et al. 2023), AusAMF provides a critical foundation for evaluating how soil fungal communities contribute to terrestrial carbon cycling and how this role may be altered under future environmental conditions. In these ways, AusAMF can inform not only ecological theory but also practical efforts in conservation planning, land management, and estimation of carbon budgets, by helping to identify regions of unique fungal diversity or areas vulnerable to functional loss.

Acknowledgements

The authors acknowledge the support of the Terrestrial Ecosystem Research Network (TERN) infrastructure, which is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS). The authors thank the Ramaciotti Centre for Genomics, S.P.U.N., and the Experiment Foundation for their support. We also thank Lindsay Lindhult for illustrations used in the figures. The majority of this work was supported by an Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA; DE220100479) awarded to A.F. and which also supported M.K.H. through a postgraduate research scholarship. J.R.P. was supported by an ARC Future Fellowship (FT190100590), S.J.W.-W. was supported by an ARC DECRA (DE210100908), as was E.E. (DE210101822). C.A.A.-T. was supported by an Academy Research Fellowship (21000058691) from the Research Council of Finland. T.V, M.Ö., and K.K. were supported by the Estonian Research Council (grants PRG1789, PRG1836) and the Centre of Excellence AgroCropFuture. J.O. was supported by the Estonian Research Council (PSG784). The authors acknowledge the Traditional Owners of the lands on which this research was conducted and pay respects to their Elders past and present. Open access publishing facilitated by Western Sydney University, as part of the Wiley - Western Sydney University agreement via the Council of Australian University Librarians.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data are available from the AusAMF database, accessible at www.ausamf.com. Raster layers used to extract environmental variables compiled into a datacube along with R code for data processing and generation of figures of this paper are available via the Figshare repository doi:10.6084/m9.figshare.28817378. Sequencing data (as fastq files) are available from the NCBI Sequence Read Archive under accession PRJNA1154677.

References

Bailey, E. 2022. "shinyBS: Twitter Bootstrap Components for Shiny." R package version 0.61.1. https://CRAN.R-project.org/package=shinyBS.

Bivand, R., and V. Gomez-Rubio. 2013. *Applied Spatial Data Analysis With R*. 2nd ed. Springer.

Camacho, C., G. Coulouris, V. Avagyan, et al. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10: 421.

Chang, W., J. Cheng, J. J. Allaire, et al. 2024. "shiny: Web Application Framework for R." CRAN, R package version 1.8.1.1. https://CRAN.R-project.org/package=shiny.

Chao, A., N. J. Gotelli, T. C. Hsieh, et al. 2014. "Rarefaction and Extrapolation With Hill Numbers: A Framework for Sampling and Estimation in Species Diversity Studies." *Ecological Monographs* 84: 45–67.

Cheng, J., B. Schloerke, B. Karambelkar, and Y. Xie. 2024. "leaflet: Create Interactive Web Maps with the JavaScript "Leaflet" Library. CRAN." R package version 2.2.2. https://CRAN.R-project.org/package=leaflet.

Davison, J., M. Moora, M. Öpik, et al. 2015. "Global Assessment of Arbuscular Mycorrhizal Fungus Diversity Reveals Very Low Endemism." *Science* 349: 970–973.

Delavaux, C. S., L. M. Smith-Ramesh, and S. E. Kuebbing. 2017. "Beyond Nutrients: A Meta-Analysis of the Diverse Effects of Arbuscular Mycorrhizal Fungi on Plants and Soils." *Ecology* 98: 2111–2119.

Delavaux, C. S., R. J. Ramos, S. L. Stürmer, and J. D. Bever. 2024. "An Updated LSU Database and Pipeline for Environmental DNA Identification of Arbuscular Mycorrhizal Fungi." *Mycorrhiza* 34: 369–373.

Delavaux, C. S., S. L. Sturmer, M. R. Wagner, U. Schütte, J. B. Morton, and J. D. Bever. 2021. "Utility of Large Subunit for Environmental Sequencing of Arbuscular Mycorrhizal Fungi: A New Reference Database and Pipeline." *New Phytologist* 229: 3048–3052.

Dumbrell, A. J., P. D. Ashton, N. Aziz, et al. 2011. "Distinct Seasonal Assemblages of Arbuscular Mycorrhizal Fungi Revealed by Massively Parallel Pyrosequencing." *New Phytologist* 190: 794–804.

Fick, S. E., and R. J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37: 4302–4315.

Flores-Moreno, H., R. L. Dalrymple, W. K. Cornwell, et al. 2023. "Is Australia Weird? A Cross-Continental Comparison of Biological, Geological and Climatological Features." *Frontiers in Ecology and Evolution* 11: 1073842.

Frew, A., J. R. Powell, M. K. Heuck, et al. 2025. "Data associated with "AusAMF: The database of arbuscular mycorrhizal fungal communities in Australia."" Figshare Digital Repository. https://doi.org/10.6084/m9.figshare.28817378.

Frew, A., P. M. Antunes, D. D. Cameron, et al. 2022. "Plant Herbivore Protection by Arbuscular Mycorrhizas: A Role for Fungal Diversity?" *New Phytologist* 233: 1022–1031.

Harris, D., R. S. Pacovsky, and E. A. Paul. 1985. "Carbon Economy of Soybean–Rhizobium–Glomus Associations." *New Phytologist* 101: 427–440.

Hawkins, H.-J., R. I. M. Cargill, M. E. Van Nuland, et al. 2023. "Mycorrhizal Mycelium as a Global Carbon Pool." *Current Biology* 33: R560–R573.

Hijmans, R. J. 2023. "raster: Geographic Data Analysis and Modeling. CRAN." R package version 3.6-26. https://CRAN.R-project.org/package=raster.

Hijmans, R. J., R. Bivand, K. Forner, J. Ooms, E. Pebesma, and M. D. Sumner. 2022. "Package 'terra'." Maintainer: Vienna, Austria. 384.

Hsieh, T. C., K. H. Ma, and A. Chao. 2016. "iNEXT: An R Package for Rarefaction and Extrapolation of Species Diversity (Hill Numbers)." *Methods in Ecology and Evolution* 7: 1451–1456.

Kohout, P., R. Sudová, M. Janoušková, et al. 2014. "Comparison of Commonly Used Primer Sets for Evaluating Arbuscular Mycorrhizal Fungal Communities: Is There a Universal Solution?" *Soil Biology and Biochemistry* 68: 482–493.

Lambers, H., M. C. Brundrett, J. A. Raven, and S. D. Hopper. 2011. "Plant Mineral Nutrition in Ancient Landscapes: High Plant Species Diversity on Infertile Soils Is Linked to Functional Diversity for Nutritional Strategies." *Plant and Soil* 348: 7–27.

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., M. Vasar, L. S. Bullington, et al. 2018. "More Bang for the Buck? Can Arbuscular Mycorrhizal Fungal Communities Be Characterized Adequately Alongside Other Fungi Using General Fungal Primers?" *New Phytologist* 220: 971–976.

Love, M. I., W. Huber, and S. Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With DESeq2." *Genome Biology* 15: 550.

Magoč, T., and S. L Salzberg. 2011. "FLASH: fast length adjustment of short reads to improve genome assemblies." *Bioinformatics* 27, no. 21: 2957–2963. https://doi.org/10.1093/bioinformatics/btr507.

Malone, B., and R. Searle. 2021. "Soil and Landscape Grid National Soil Attribute Maps - pH - CaCl2 (3″ resolution) - Release 2. v1."

Malone, B., and R. Searle. 2024. "Soil and Landscape Grid National Soil Attribute Maps - Total Phosphorus (3″ resolution) - Release 2. v1."

Marschner, H., and B. Dell. 1994. "Nutrient Uptake in Mycorrhizal Symbiosis." *Plant and Soil* 159: 89–102.

Massicotte, P., and A. South. 2023. "rnaturalearth: World Map Data from Natural Earth." CRAN, R package version 1.0.1. https://CRAN.R-project.org/package=rnaturalearth.

McKenzie, N. N., D. D. Jacquier, R. R. Isbell, and K. K. Brown. 2004. *Australian Soils and Landscapes: An Illustrated Compendium*. CSIRO publishing.

McMurdie, P. J., and S. Holmes. 2013. "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." *PLoS One* 8: e61217.

Neuenkamp, L., M. Moora, M. Öpik, et al. 2018. "The Role of Plant Mycorrhizal Type and Status in Modulating the Relationship Between Plant and Arbuscular Mycorrhizal Fungal Communities." *New Phytologist* 220: 1236–1247.

Oksanen, J., F. G. Blanchet, P. Legendre, et al. 2015. "vegan: Community Ecology Package." http://CRAN.R-project.org/package=vegan.

Öpik, M., A. Vanatoa, E. Vanatoa, et al. 2010. "The Online Database MaarjAM Reveals Global and Ecosystemic Distribution Patterns in Arbuscular Mycorrhizal Fungi (Glomeromycota)." *New Phytologist* 188: 223–241.

Öpik, M., J. Davison, M. Moora, and M. Zobel. 2014. "DNA-Based Detection and Identification of Glomeromycota: The Virtual Taxonomy of Environmental Sequences." *Botany* 92: 135–147.

Orians, G. H., and A. V. Milewski. 2007. "Ecology of Australia: The Effects of Nutrient-Poor Soils and Intense Fires." *Biological Reviews* 82: 393–423.

Pagès, H., P. Aboyoun, R. Gentleman, and S. DebRoy. 2024. "Biostrings: Efficient manipulation of biological strings." CRAN, R package version 2.70.3. https://bioconductor.org/packages/Biostrings.

Paradis, E., and K. Schliep. 2019. "Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." *Bioinformatics* 35: 526–528.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Read, D. J., and J. Perez-Moreno. 2003. "Mycorrhizas and Nutrient Cycling in Ecosystems – A Journey Towards Relevance?" *New Phytologist* 157: 475–492.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. "VSEARCH: A Versatile Open Source Tool for Metagenomics." *PeerJ* 4: e2584.

Rossel, R. V., C. Chen, M. Grundy, et al. 2014. "Soil and Landscape Grid National Soil Attribute Maps - Total Nitrogen (3″ resolution) - Release 1. v5."

Schoch, C. L., K. A. Seifert, S. Huhndorf, et al. 2012. "Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi." *Proceedings of the National Academy of Sciences of the United States of America* 109: 6241–6246.

Sievert, C., J. Cheng, and G. Aden-Buie. 2024. "bslib: Custom "Bootstrap" "Sass" Themes for "shiny" and "rmarkdown"." CRAN, R package version 0.7.0. https://CRAN.R-project.org/package=bslib.

Smith, S. E., and D. J. Read. 2008. *Mycorrhizal Symbiosis*. Academic Press.

Stern, H., G. De Hoedt, and J. Ernst. 2000. "Objective Classification of Australian Climates." *Australian Meteorological Magazine* 49: 87–96.

Stockinger, H., M. Krüger, and A. Schüßler. 2010. "DNA Barcoding of Arbuscular Mycorrhizal Fungi." *New Phytologist* 187: 461–474.

Tedersoo, L., M. Bahram, L. Zinger, et al. 2022. "Best Practices in Metabarcoding of Fungi: From Experimental Design to Results." *Molecular Ecology* 31: 2769–2795.

Terrestrial Ecosystem Research Network. 2024. "Soil and Landscape Grid of Australia."

Teste, F. P., M. D. Jones, and I. A. Dickie. 2020. "Dual-Mycorrhizal Plants: Their Ecology and Relevance." *New Phytologist* 225: 1835–1851.

Thiéry, O., M. Vasar, T. Jairus, et al. 2016. "Sequence Variation in Nuclear Ribosomal Small Subunit, Internal Transcribed Spacer and Large Subunit Regions of Rhizophagus Irregularis and Gigaspora Margarita Is High and Isolate-Dependent." *Molecular Ecology* 25: 2816–2832.

Vasar, M., J. Davison, L. Neuenkamp, et al. 2021. "User-Friendly Bioinformatics Pipeline gDAT (Graphical Downstream Analysis Tool) for Analysing rDNA Sequences." *Molecular Ecology Resources* 21: 1380–1392.

Větrovský, T., Z. Kolaříková, C. Lepinay, et al. 2023. "GlobalAMFungi: A Global Database of Arbuscular Mycorrhizal Fungal Occurrences From High-Throughput Sequencing Metabarcoding Studies." *New Phytologist* 240: 2151–2163.

Wadoux, A., M. R. Dobarco, B. Malone, B. Minasny, A. McBratney, and R. Searle. 2022. Soil and Landscape Grid National Soil Attribute Maps - Organic Carbon (1″ resolution) - Release 1. v3.

White, A., B. Sparrow, L. Emrys, et al. 2012. *AusPlots Rangelands Survey Protocols Manual, 1.2.9*. University of Adelaide Press.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Wickham, H., J. Hester, and J. Bryan. 2024. "readr: Read Rectangular Text Data." CRAN, R package version 2.1.5. https://CRAN.R-project.org/package=readr.

Wickham, H., R. François, L. Henry, K. Müller, and D. Vaughan. 2023. "dplyr: A Grammar of Data Manipulation." CRAN, R package version 1.1.4. https://CRAN.R-project.org/package=dplyr.

Xie, Y., J. Cheng, and X. Tan. 2024. "DT: A Wrapper of the JavaScript Library "DataTables"." CRAN, R package version 0.33. https://CRAN.R-project.org/package=DT.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam. 2017. "Ggtree: An r Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data." *Methods in Ecology and Evolution* 8: 28–36.

Zund, P. 2022. "Soil and Landscape Grid National Soil Attribute Maps - Available Phosphorus (3″ resolution) - Release 1. v1."