GlobalAMFungi : a global database of arbuscular mycorrhizal fungal occurrences from high-throughput sequencing metabarcoding studies

Tomáš Větrovský, Zuzana Kolaříková, Clémentine Lepinay, Sandra Awokunle Hollá, John Davison, Anna Fleyberková, Anastasiia Gromyko, Barbora Jelínková, Miroslav Kolařík, Manuela Krüger, et al.

New Phytologist

Research

Methods

GlobalAMFungi: a global database of arbuscular mycorrhizal fungal occurrences from high-throughput sequencing metabarcoding studies

Tomáš Větrovský[1], Zuzana Kolaříková[2], Clémentine Lepinay[1], Sandra Awokunle Hollá[1], John Davison[3], Anna Fleyberková[1], Anastasiia Gromyko[1], Barbora Jelínková[1], Miroslav Kolařík[1], Manuela Krüger[2], Renata Lejsková[1], Lenka Michalčíková[1], Tereza Michalová[1], Mari Moora[3], Andrea Moravcová[1,4], Štěpánka Moulíková[1], Iñaki Odriozola[1], Maarja Öpik[3], Monika Pappová[1], Sarah Piché-Choquette[1], Jakub Skřivánek[1,4], Lukáš Vlk[1], Martin Zobel[3], Petr Baldrian[1] and Petr Kohout[1,4]

[1]Institute of Microbiology of the Czech Academy of Sciences, Vídeňská 1083, 142 20, Prague, Czechia; [2]Institute of Botany of the Czech Academy of Sciences, Zámek 1, 252 43, Průhonice, Czechia; [3]Institute of Ecology and Earth Sciences, University of Tartu, J. Liivi St 2, 504 09, Tartu, Estonia; [4]Faculty of Science, Charles University, Albertov 6, 128 43, Prague, Czechia

Author for correspondence:
Petr Kohout
Email: petr.kohout@natur.cuni.cz

Received: *11 April 2023*
Accepted: *4 September 2023*

New Phytologist (2023) **240:** 2151–2163
doi: 10.1111/nph.19283

Key words: arbuscular mycorrhizal fungi, fungal biogeography, fungal diversity, glomeromycota, glomeromycotina, high-throughput sequencing.

Summary

- Arbuscular mycorrhizal (AM) fungi are crucial mutualistic symbionts of the majority of plant species, with essential roles in plant nutrient uptake and stress mitigation. The importance of AM fungi in ecosystems contrasts with our limited understanding of the patterns of AM fungal biogeography and the environmental factors that drive those patterns.
- This article presents a release of a newly developed global AM fungal dataset (GlobalAMFungi database, https://globalamfungi.com) that aims to reduce this knowledge gap. It contains almost 50 million observations of Glomeromycotinian AM fungal amplicon DNA sequences across almost 8500 samples with geographical locations and additional metadata obtained from 100 original studies.
- The GlobalAMFungi database is built on sequencing data originating from AM fungal taxon barcoding regions in: i) the small subunit rRNA (SSU) gene; ii) the internal transcribed spacer 2 (ITS2) region; and iii) the large subunit rRNA (LSU) gene.
- The GlobalAMFungi database is an open source and open access initiative that compiles the most comprehensive atlas of AM fungal distribution. It is designed as a permanent effort that will be continuously updated by its creators and through the collaboration of the scientific community. This study also documented applicability of the dataset to better understand ecology of AM fungal taxa.

Introduction

Arbuscular mycorrhizal (AM) fungi, belonging to the phylum Mucoromycota (Spatafora *et al.*, 2016), represent crucial plant mutualistic symbionts interacting with >70% of plant species, including many agricultural crops (Brundrett & Tedersoo, 2018). This makes them essential for the establishment and productivity of plants in nutrient-limited soils and for mitigation of plant stress (e.g. Smith & Read, 2008; Kakouridis *et al.*, 2022). There are currently 332 described and many more unknown species of AM fungi (Wijayawardene *et al.*, 2020), which differ in their beneficial effects on host plants (Kiers *et al.*, 2011; Koch *et al.*, 2017) and contribution to a range of other ecosystem processes, such as soil aggregation and nutrient retention (Chagnon *et al.*, 2013; Qiu *et al.*, 2022).

Therefore, it is of considerable importance to determine how environmental factors affect the diversity and distribution of AM fungi. The advent of high-throughput sequencing-powered metabarcoding has enabled researchers to explore the global distribution of fungi, using fungal-specific primers, targeting the internal transcribed spacer (ITS) region of rRNA, which is recognized as a common region for barcoding of fungal taxa (Schoch *et al.*, 2012). Although there have been several attempts to investigate global fungal diversity and distribution (Tedersoo *et al.*, 2014, 2022; Maestre *et al.*, 2015; Větrovský *et al.*, 2019), none of these initiatives provided a thorough description of AM

fungal communities, because of several reasons. First, PCR primers designed for amplification of the ITS region of Dikarya (Ascomycota and Basidiomycota) are suboptimal (often include several mismatches) for amplification of AM fungi (Kohout *et al.*, 2014; Lekberg *et al.*, 2018). Second, shallow sequencing of fungal communities has been used, resulting in low coverage of AM fungal taxa potentially due to their smaller DNA share in soil DNA pool. Third, there is no single and commonly used barcode region for delineating AM fungal taxa that is short enough for sequencing using second-generation high-throughput sequencing methods.

To overcome these limitations, specific primer systems have been developed for targeted analysis of AM fungal communities. These often include nested-PCR approach, with AM fungal-specific primers applied in at least one of the steps and target three different barcoding regions in genes coding for rRNA (Kohout *et al.*, 2014; Van Geel *et al.*, 2014; Kolaříková *et al.*, 2021), part of the small subunit rRNA (SSU) gene, the internal transcribed spacer (ITS) region, and part of the large subunit rRNA (LSU) gene. These barcoding regions have been employed in metabarcoding studies focussed on AM fungal communities for more than a decade and each with its advantages and disadvantages (Öpik *et al.*, 2010; Nilsson *et al.*, 2019a).

The majority of studies aiming to describe AM fungal community composition have so far relied on sequencing the SSU rRNA region (Öpik *et al.*, 2014), although the other two barcoding regions have also been repeatedly utilized (see Delavaux *et al.*, 2022). Due to the obligatory symbiotic lifestyle, only a subset of AM fungal species is available in pure culture collections (e.g. INVAM (Stürmer *et al.*, 2021); BEG, https://www.i-beg.eu/). Therefore, formally described AM fungal species represent just a small fraction of the AM fungal molecular diversity characterized using environmental DNA sequences (Ohsowski *et al.*, 2014; Öpik *et al.*, 2014). Moreover, AM fungal phylogroups that are identified only on the basis of their DNA are categorized differently by studies using different AM fungal barcoding regions, which limits the direct comparability of such data.

Despite the ecological importance of AM fungi, the global distribution of AM fungal taxa has not yet been adequately described (e.g. Davison *et al.*, 2015; Stürmer *et al.*, 2018). This is mostly due to large sampling biases towards heavily studied regions in Europe (e.g. Davison *et al.*, 2012; Kohout *et al.*, 2015; Clavel *et al.*, 2021), North America (e.g. Lekberg *et al.*, 2013; House & Bever, 2020; Rodriguez-Ramos *et al.*, 2021) and China (e.g. Xiang *et al.*, 2014; Jiang *et al.*, 2018; Fan *et al.*, 2020), while many tropical regions or large parts of Asia remain understudied. Besides that, only a handful of studies has investigated the composition of AM fungal communities using metabarcoding at a global scale (Öpik *et al.*, 2013; Davison *et al.*, 2015, 2021; Vasar *et al.*, 2021), all of which used a sparse sampling regime.

A previous attempt to collect and validate published data on the composition of complete fungal communities was solely focussed on metabarcoding studies using general fungal primers, targeting the ITS region of rRNA (Větrovský *et al.*, 2020). Although the resulting freely available GlobalFungi database, which in its current version includes > 57 000 samples from > 500 fungal metabarcoding studies, has proven to be a valuable tool for determining the biogeography of various fungal taxa (Réblová *et al.*, 2021; Sugiyama & Sato, 2021; Mozzachiodi *et al.*, 2022; Moulíková *et al.*, 2023), for identifying non-native symbiotic fungi associating with alien plants (Vlk *et al.*, 2020a,b) or for estimating global fungal species richness (Baldrian *et al.*, 2022b), it does not sufficiently represent the distribution and community composition of AM fungal taxa due to the limitations of general fungal primer systems. As a consequence, only *c.* 0.29% of the sequences contained in the GlobalFungi are assigned to the Glomeromycotina (https://globalfungi.com, Release 4).

To overcome current limitations of the understanding of AM fungal community composition at the global scale, we have now compiled a comprehensive collection of published data on the composition of Glomeromycotinian AM fungal communities described by at least one of the three most commonly used molecular markers (SSU, ITS and/or LSU) derived from soil or plant root samples collected in terrestrial environments. This enabled us to construct the most complete public dataset currently available containing information about the diversity of AM fungal communities and the distribution of AM fungal species or molecular taxa: the novel GlobalAMFungi database (https://globalamfungi.com). To document the potential applicability of the GlobalAMFungi database, we used the collated dataset to address preferences of AM fungal families to different biomes and substrate types (roots vs soil). As is the case for its sister GlobalFungi database, this database is designed as a permanent effort that will be continuously updated by its creators in collaboration with members of the scientific community, and it will represent a permanent source of FAIR (Findable, Accessible, Interoperable, Reusable) data on the distribution of AM fungi.

Materials and Methods

Studies selection

We explored studies aiming to characterize the composition of AM fungal communities using metabarcoding of the most commonly used barcoding regions. While the ITS region of rRNA is widely considered to be a suitable molecular marker for describing general fungal communities, other regions have typically been used in AM fungal metabarcoding studies. The most widely used markers have been: the variable regions V4 and V5 of the small subunit (SSU or 18S) of the rRNA gene; the internal transcribed spacer (ITS); and the variable regions D1 and D2 of the large subunit (LSU or 28S) of the rRNA gene (Öpik & Davison, 2016). To make our database comprehensive, we decided to include all metabarcoding studies using AM fungal-specific primers targeting one of the three above-mentioned barcoding regions, published up to 2021.

A publication search was performed in May 2021 using the following search criteria: 'arbuscular AND 454; arbuscular AND Illumina; arbuscular AND pyroseq*; arbuscular AND PacBio; arbuscular AND Ion Torrent; glomero* AND 454; glomero* AND Illumina; glomero* AND pyroseq*; glomero* AND

PacBio; glomero* AND Ion Torrent; arbuscular AND community AND sequencing' on Scopus, Web of Science, and Google Scholar. The search was limited to publications where the main text was written in English. This search returned *c.* 500 publications. Studies, and consequently samples, were only included in the final dataset where the following criteria were met: (1) the environmental DNA was subject to high-throughput amplicon sequencing using AM fungal-specific primers targeting SSU, ITS, or LSU barcoding regions; (2) samples originated from soil or plant roots and were not subject to experimental treatments that artificially modified environmental conditions at the sampled sites (e.g. experimental manipulation of temperature, precipitation, or nutrient availability in field experiments, glasshouse controlled conditions, and AM fungal inoculation experiments); (3) the precise geographic location of each sample was available; (4) sequencing data (either in FASTA with PHRED scores reported or FASTQ format) were publicly available or provided by the authors of the study upon request; and (5) the sequences were unambiguously assigned to samples. In total, 100 publications contained samples and sequences that matched our criteria.

High-throughput sequencing data processing

Raw reads from 100 studies, covering 8464 individual samples, were filtered and trimmed using SEED software v.2.1 (Větrovský *et al.*, 2018). Sequences with a mean PHRED quality score of < 20 were discarded. The SEED tool was also employed to search for the sequence motives, allowing to set required number of mismatches. Each sequence was then labelled by the combination of the sample ID and sequence ID. Sequences were divided according to the focal barcoding region into three datasets (SSU, ITS, and LSU) and then processed independently, that is assigned to molecular and/or morphological taxa specific to one of the three regions studied. In case of SSU and ITS regions, the sequences were assigned to virtual taxa (VT) and species hypothesis (SH), respectively. These two similar concepts represent grouping of arbuscular mycorrhizal fungal (in case of VT) and fungal taxa in general (in case of SH) that are defined based on genetic similarity rather than physical characteristics (Kõljalg *et al.*, 2013; Öpik *et al.*, 2014). Neither the VT nor the SH concept is intended to replace the Linnaean classification system. On the contrary, there is no molecular-based sorting of AM fungal taxa for the LSU sequences. Therefore, the obtained LSU AM fungal sequences were classified according to the closest match among AM fungal OTUs in the curated and most comprehensive reference database of AM fungal LSU sequences (Delavaux *et al.*, 2021, 2022).

SSU dataset Altogether, we collected sequencing data and metadata from 75 studies that used the SSU marker for AM fungal community metabarcoding and fulfilled our selection criteria. Although all authors of the identified papers used the V4 and V5 variable regions of SSU to identify AM fungal taxa, the studies differed considerably in the applied primer combinations and sequencing techniques, which resulted in insufficient overlap of forward and reverse reads of generated sequences among all studies. To overcome this limitation, we decided to build the SSU dataset of the GlobalAMFungi database on the V4 variable region of SSU (Fig. 1). All sequences were subject to a search of the sequence motif corresponding to the AMV4.5NF primer (Sato *et al.*, 2005) with two allowed mismatches. The two mismatches were allowed to avoid potential loss of some AM fungal taxa, because of their slight genetic variability in this region. All sequences were then trimmed to 190 bases downstream of the searched motif (region common to all selected individual datasets). Importantly, the V4 region is characterized by the highest sequence dissimilarity among AM fungal sequences (Vasar *et al.*, 2017) and therefore represents a suitable barcoding region for short reads obtained from the most widely used high-throughput sequencing methods. The resulting SSU sequences were classified according to the closest match among representative sequences of MaarjAM VT (Öpik *et al.*, 2010) using BLASTN (Altschul *et al.*, 1997). We used the most recent release of type sequences of VT from the MaarjAM database from 5 June 2019, with a slight correction of some Latin names of the AM fungal species according to the species list in http://www.amf-phylogeny.com/ in order to follow up to date taxonomy and nomenclature. A sequence was classified to one of the AM fungal VT only when the following criteria were met: sequence coverage ≥ 98% and sequence similarity ≥ 97%. All other sequences with coverage ≥ 95% and similarity ≥ 90% to any VT were considered

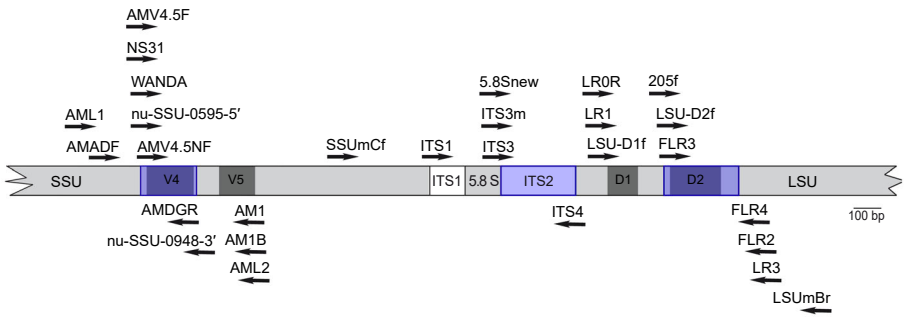

Fig. 1 Ribosome-encoding gene operon with the three target barcoding regions highlighted in blue. The barcoding region in the small subunit rRNA gene (SSU) covers the V4 variable region; the barcoding region in the internal transcribed spacer (ITS) region covers the ITS2 spacer; and the barcoding region in the large subunit rRNA gene (LSU) covers the D2 variable region. All PCR primers used for the amplification of one of the three barcoding regions from the 100 surveyed studies are shown in their position along the rRNA operon. For more information, see Supporting Information Table S1, which contains all information about primers used in the original publications.

as AM fungi, and they are accessible through a BLAST search in the GlobalAMFungi database. The number of unique sequences (termed 'sequence variants') accessible in the SSU part of the GlobalAMFungi database is 14 337 166.

ITS dataset Altogether, we collated sequencing data and metadata from eight ITS-based AM fungal community metabarcoding studies that fulfilled our selection criteria. Among these eight studies, two sequenced amplicons covering the full ITS region (including ITS1, 5.8S, and ITS2) of rRNA, while in the remaining six studies only the ITS2 region was sequenced. All eight studies applied AM fungal-specific primers for metabarcoding the AM fungal communities (Supporting Information Table S1; Fig. 1). Considering that only two identified studies (with relatively low numbers of samples) used both ITS1 and ITS2, we built the ITS part of the GlobalAMFungi database on the ITS2 region only. ITS regions were extracted using the PERL script ITSx v.1.0.11 (Bengtsson-Palme *et al.*, 2013). Extracted ITS2 sequences were classified according to the closest match among representative sequences of SH (Kõljalg *et al.*, 2013) from the dynamic version of the UNITE database (BLASTDBv.5, general release 9.0 from 16 October 2022) using BLASTN. A sequence was assigned to an AM fungal SH when the following thresholds were met: sequence coverage ≥ 98% and similarity ≥ 97%. All other sequences with a best hit to an AM fungal SH were considered as potential AM fungi and are accessible through a BLAST search in the GlobalAMFungi database. All sequences that were not assigned to an AM fungal SH or sequences with a closest BLAST hit to non-AM fungal SH were discarded. The number of sequence variants accessible in the ITS part of the GlobalAMFungi database is 2563 095.

LSU dataset Altogether, we collated sequencing data and metadata from 17 studies that fulfilled our selection criteria and used the LSU marker for AM fungal community metabarcoding. Among these 17 studies, six sequenced both the D1 and D2 regions of the LSU, while in the remaining 11 studies, only the D2 region was subject to amplicon sequencing (Table S1). Besides this, the AM fungal-specific primers used for amplification of even the same barcoding region of LSU varied markedly among the selected studies (Fig. 1). To overcome the problem of comparing sequences with highly variable lengths, we built the LSU part of the GlobalAMFungi database based on the D2 region only. Specifically, we defined the barcoding region for the LSU part of the GlobalAMFungi database as a DNA sequence between the binding sites of the FLR3 and FLR4 primers, with an approximate length of 360 bases. Two mismatches in each of the searched primer sequences were allowed. All flanking regions were trimmed from the sequences. Extracted sequences covering the variable D2 region of the LSU were classified according to their closest match among AM fungal OTUs in the curated and most comprehensive reference database of AM fungal LSU sequences (Delavaux *et al.*, 2021, 2022) using BLASTN. A sequence was assigned to an AM fungal OTU only when: sequence coverage ≥ 98% and sequence similarity ≥ 97% (Krüger *et al.*, 2015). All classified sequences as well as all other sequences with a best hit to an AM fungal LSU sequence were retained and are accessible through a BLAST search in the GlobalAMFungi database. The number of sequence variants in the LSU part of the GlobalAMFungi database is 1203 288.

Sample metadata

Similar to the GlobalFungi database (Větrovský *et al.*, 2020), we collected a comprehensive set of sample metadata from the published papers, public repositories, or through direct contact with the authors of individual studies. Samples were assigned to continents, and all sites were categorized into biomes following the classification of the Environment Ontology (http://www.ontobee.org/ontology/ENVO). In addition to the metadata provided by the authors of each study, we also extracted bioclimatic variables, mean annual temperature (MAT), and mean annual precipitation (MAP) from the global CHELSA database (Karger *et al.*, 2017) for each sample based on its geographic coordinates. Similarly, soil chemistry characteristics were extracted from the SoilGrids database (Hengl *et al.*, 2017). Because pH was found to be the most relevant soil variable in determining the distribution of AM fungal taxa (Davison *et al.*, 2021) – and since the majority of studies focussed on soil AM fungal communities in the top soil layers – we extracted pH values for the first 5 cm of the soil horizon from the SoilGrids database (250 m resolution level). We also provide information about host plant species (if available in the original publication or on request from the authors) for samples originating from plant roots. The complete list of metadata included in the GlobalAMFungi database is presented in Table S2. Geographic distribution of the samples is presented in Fig. S1.

Database design

GlobalAMFungi was written in R using package SHINY deployed by SHINYPROXY (v.2.1.0). Data are stored in the open source relational database MariaDB (v.10.8.3). The graphical user interface is accessible through all common browsers and best viewed using Firefox, Chrome, or Edge.

Statistical analysis

To document one of the potential applicability of the GlobalAMFungi database, we addressed the preferences of AM fungal families to different biomes and substrate types (roots vs soil). To test the null hypothesis of no differences in relative abundances of AM fungal families between substrate types (categorical variables with two levels: root and soil) and biomes (categorical variable with five levels: cropland, forest, grassland, shrubland, and woodland), we used linear models. The raw sequence counts of each AM fungal family were centred log ratio (CLR) transformed using the function transform of the R package MICROBIOME (Lahti & Shetty, 2018) to account for the compositionality of the data and used as response variables in linear models. In both models (to test the preferences of AM fungal families to different biomes and substrate types), we used the dataset based on the amplification of the SSU

Table 1 Data structure in the GlobalAMFungi database.

			Sequences[a]		Glomeromycotan taxonomy			
	Studies	Samples	All	Glomeromycotina	Molecular taxa[b]	Species	Genera	Families
SSU	75	6912	74 087 884	41 879 402	338 VT	30	13	9
ITS2	8	883	9677 697	6435 286	1033 SH	61	45	15
LSU	17	669	5251 848	1519 747	58 taxa	38	23	9

[a]The numbers listed in the 'Glomeromycotina' represent sequences, which were assigned to existing AMF VT (in case of SSU), AMF SH (in case of ITS), and AMF taxa (in case of LSU). The numbers listed as 'All' represent all high-quality sequences regardless of their taxonomical assignment.
[b]Molecular taxa were defined differently among the three selected barcoding regions. Concept of virtual taxa (VT), based on Öpik *et al*. (2010), was used for sorting of SSU sequences into arbuscular mycorrhizal fungal taxonomical units. Concept of species hypothesis (SH), based on Kõljalg *et al*. (2013), was used for sorting of ITS2 sequences into arbuscular mycorrhizal fungal taxonomical units. For sorting of LSU sequences into arbuscular mycorrhizal fungal taxonomical units, we used representative sequences from the LSU database of arbuscular mycorrhizal taxa published by Delavaux *et al*. (2021, 2022).

target gene. Homoscedasticity and normal distribution of residuals were verified with visual inspection of residual plots. In all linear models, significant differences among groups were assessed using Tuckey's HSD *post hoc* test.

To assess whether different target genes (ITS2, LSU, and SSU) tended to over or underrepresent different AM fungal families, we first transformed the raw sequence counts to relative abundances and then computed mean and 95% confidence intervals through bootstrap resampling with replacement (with 1000 iterations), using the functions boot and boot.ci of the R package BOOT (Canty & Ripley, 2022). Relative sequence counts assigned to each AM fungal family when using different target genes were biased by the fact that the reference databases (MaarjAM for SSU sequences, UNITE for ITS2 sequences, and reference datasets published by Delavaux *et al*., 2021, 2022 for LSU sequences) themselves differed on the representation of different families (i.e. the number of AM fungal taxa belonging to each family differed between different target genes). To account for this bias, we computed the relative number of AM fungal taxa within each family for each target gene and then subtracted that value to above-computed bootstrap mean and confidence intervals. If the difference resulted in negative values with confidence intervals not overlapping zero, the family was considered underrepresented by the target gene in the GlobalAMFungi database, as compared to the reference database (e.g. MaarjAM and UNITE); in contrast, if the difference resulted in positive confidence intervals, the family was considered over-represented by the target gene; lastly, if confidence intervals included zero, the representation of the family was considered not different from the null expectation.

Results

The SSU dataset

Altogether, we collated 6912 samples with data from the focal SSU barcoding region among the 75 selected studies. In total, 41 879 402 sequences were assigned to 338 AM fungal VT (Table 1). The number of sequences assigned to VT, ranged from 1 to 363 607 per sample, with an average value of 6277. A total of 289 VT were recorded in ≥ 30 samples, while 27 VT were found in fewer than 10 samples. Sequences that were assigned to VT from the family Glomeraceae represented 85% of the whole SSU dataset of the GlobalAMFungi database (Fig. 2), followed by Claroideoglomeraceae (6.1%), Acaulosporaceae (2.2%), Paraglomeraceae (2.1%), Diversisporaceae (2%), and Gigasporaceae (1.8%). While Glomeraceae and Claroideoglomeraceae were over-represented than expected by chance in the SSU dataset compared with their representation in the reference database, all other identified AM fungal families were underrepresented (Fig. S2). The Glomeraceae were the most diverse AM fungal family with 238 VT in total.

The majority of samples in the SSU dataset of the GlobalAMFungi database originated from North America and Europe, while South America was the least represented continent (Fig. 3). Among countries, the largest share of samples was derived from Canada (15%), the USA (14%), and China (13%). In total, the SSU dataset contained samples from 100 countries, but almost

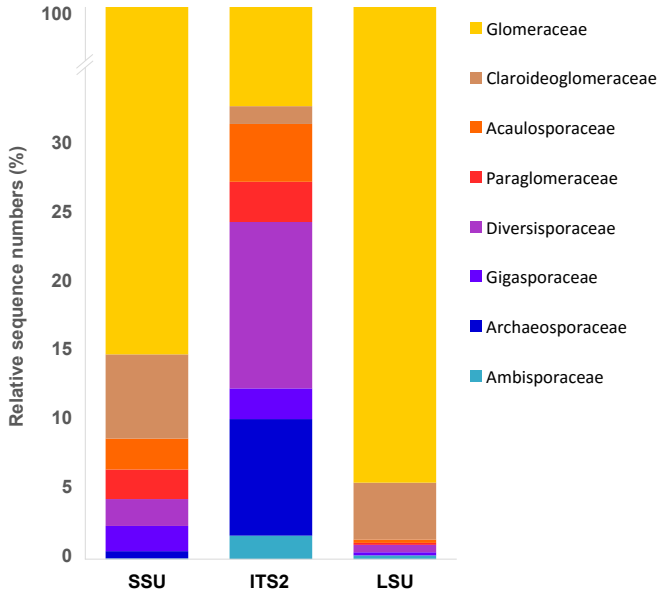

Fig. 2 Relative sequence numbers of fungal taxa belonging to the most common arbuscular mycorrhizal fungal families among the three datasets of the GlobalAMFungi database.

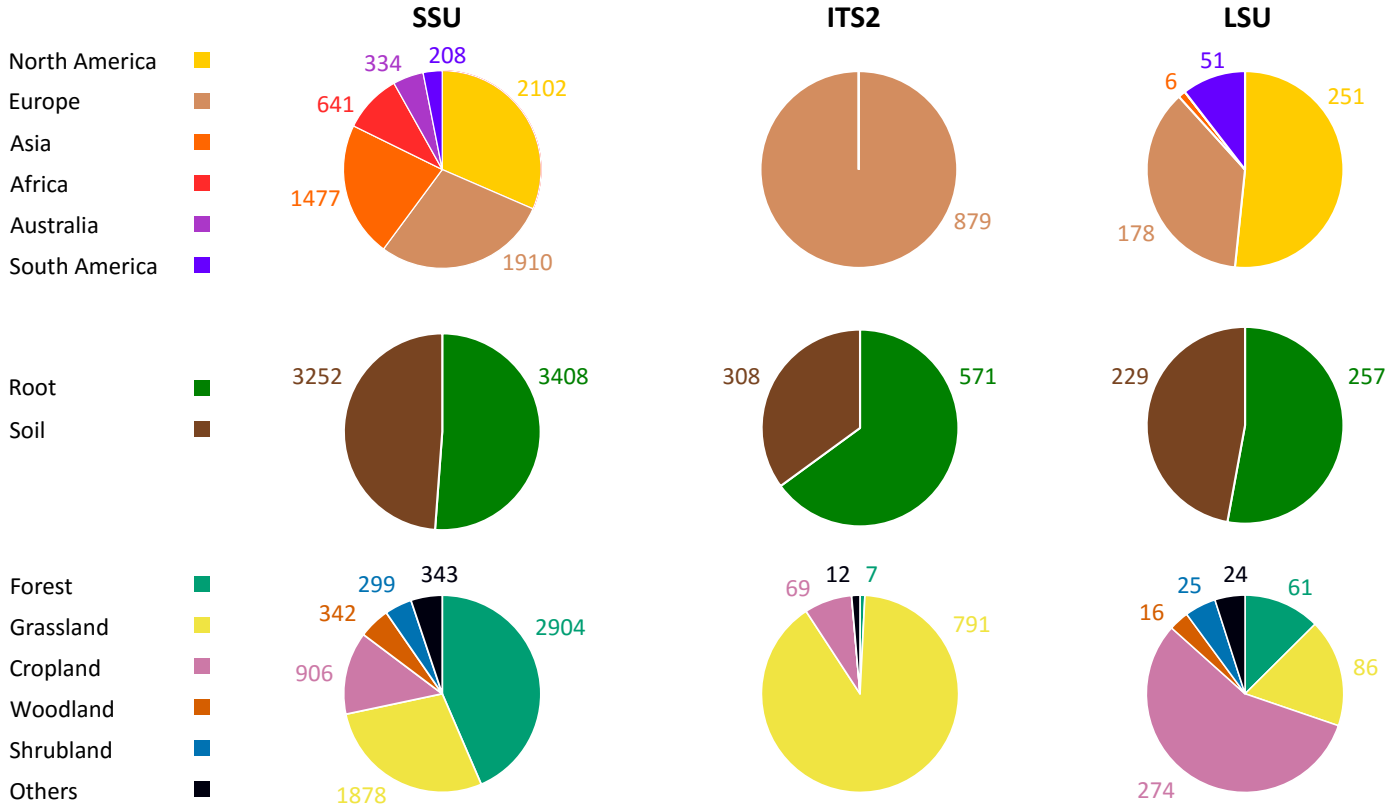

Fig. 3 Distribution of the samples among geographical regions (top row), sample sources (middle row), and biomes (bottom row) within the three datasets of the GlobalAMFungi database. Numbers represent sample counts.

half of the countries were represented by one or two samples only. The samples included in this dataset covered a wide range of environmental conditions: pH (2.5–9.0), mean annual temperature (−15°C to +29°C), and mean annual precipitation (20–9700 mm). Biomes were predominantly represented by forest and grassland in the SSU dataset, followed by cropland, woodland, and shrubland (Fig. 3). The less sampled biomes included tundra and desert. Sample origin was almost equally split between soil and plant roots (Fig. 3).

The distribution of VT across continents indicated a high prevalence of ubiquitous AM fungal taxa. While 147 VT were recorded from all continents, only 31 VT were identified in samples originating from a single continent (Fig. 4). The highest VT richness (316 VT) was recorded from Eurasia, which likely reflects the biased sampling effort, because more than half of the samples in the SSU part of the GlobalAMFungi database originated from this region. Interestingly, the second most VT-rich geographical region was Africa with 280 recorded VT, although < 8% of the samples in our database originated from there.

The relative abundance of VT belonging to different AM fungal families differed among the biomes represented in our database (Fig. 5). While Glomeraceae, Claroideoglomeraceae, and Paraglomeraceae showed a higher preference for biomes dominated by herbaceous and graminoid vegetation (cropland and grassland), compared with forest, shrubland, or woodland, distribution of Ambisporaceae and Pacisporaceae followed the exact

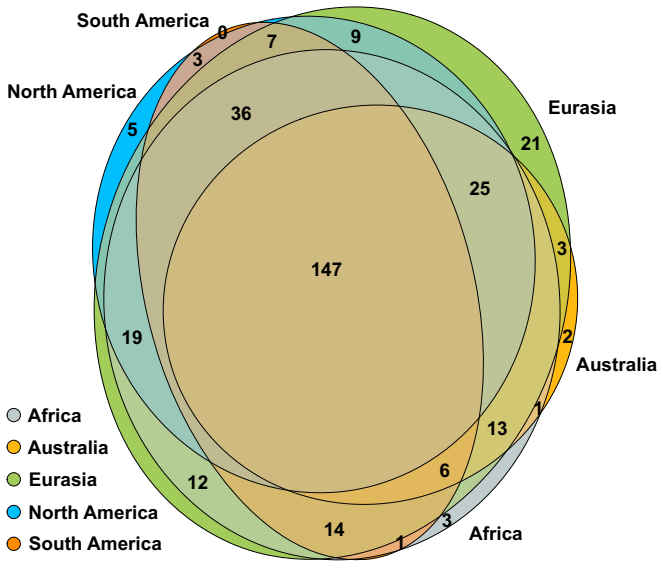

Fig. 4 Venn diagram visualizing the number of arbuscular mycorrhizal fungal virtual taxa in the small subunit rRNA dataset of the GlobalAMFungi database that are unique to or shared between different geographical regions.

opposite pattern (Fig. S3). Among the other AM fungi families, the Archaeosporaceae were the least represented in croplands, while Acaulosporaceae reached the highest relative abundances in

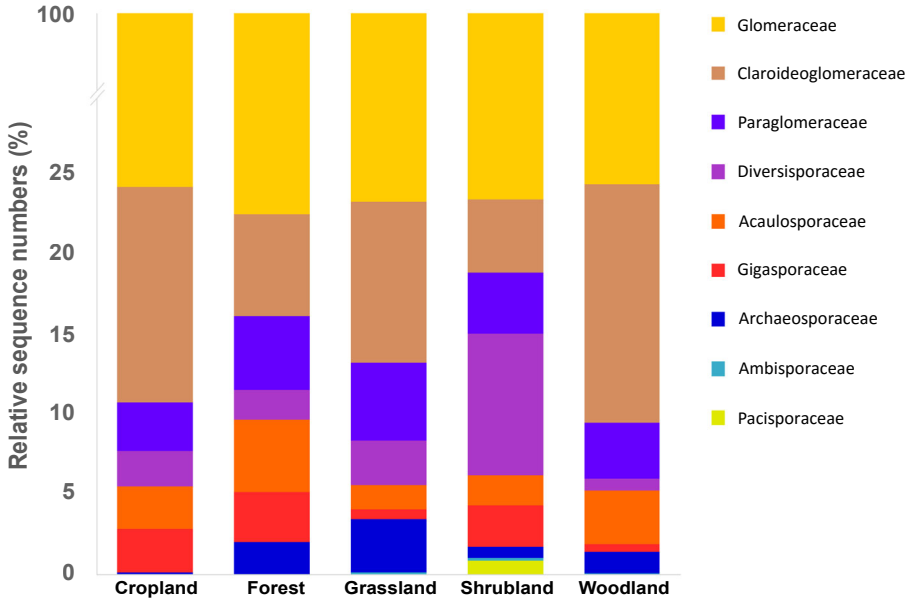

Fig. 5 Relative sequence numbers of arbuscular mycorrhizal fungal virtual taxa belonging to different families among the five most studied biomes within the small subunit rRNA dataset of the GlobalAMFungi database.

forests. In all biomes, Glomeraceae accounted for > 75% of AM fungal sequences.

The relative abundance of VT belonging to different AM fungal families also differed between soil and plant roots (Fig. 6). While Glomeraceae, Ambisporaceae, and Pacisporaceae showed higher relative share in root compared with soil samples, distribution of Claroideoglomeraceae, Paraglomeraceae, Diversisporaceae, and Archaeosporaceae followed the exact opposite pattern (Fig. S4). In case of the most abundant AM fungal family, the Glomeraceae accounted for > 85% of sequences from root samples, while < 70% of sequences were assigned to Glomeraceae in soil samples.

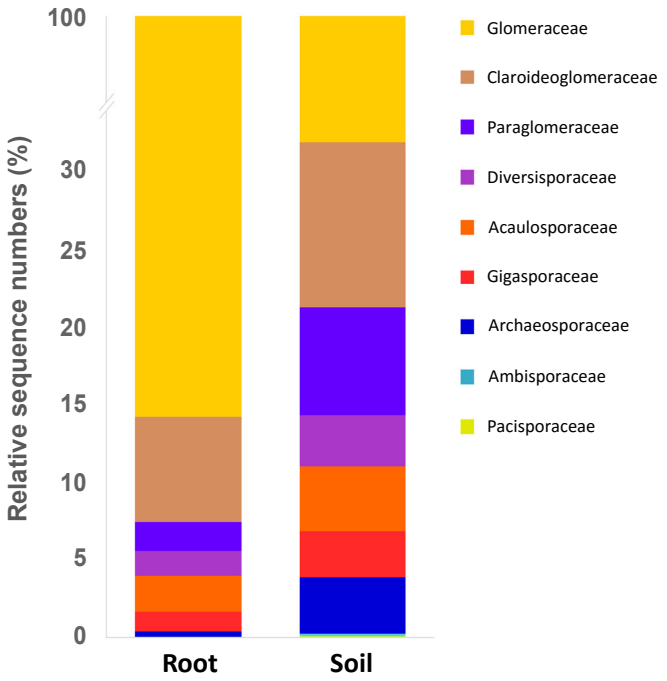

Fig. 6 Relative sequence numbers of fungal virtual taxa belonging to the most common arbuscular mycorrhizal fungal families in the two sample sources (roots and soil). Data are shown for the small subunit rRNA dataset within the GlobalAMFungi database.

The ITS dataset

We collated 883 samples with data from the ITS2 among the eight selected studies. In total, 6435 286 sequences were assigned to 1033 AM fungal species hypotheses (SH; Table 1). The number of sequences assigned to SH, ranged from 1 to 90 388 per sample, with an average value of 7247. A total of 371 SH were recorded in ≥ 20 samples, while 501 SH were found in fewer than 10 samples. Among the phylogenetic lineages (Fig. 2), sequences that were assigned to SH from the family Glomeraceae represented 67% of the whole ITS2 dataset of the GlobalAMFungi database, followed by Diversisporaceae (12%), Archaeosporaceae (8.4%), Acaulosporaceae (4.2%), Paraglomeraceae (2.9%), and Gigasporaceae (2.2%). While Glomeraceae and Paraglomeraceae were over-represented than expected by chance in the ITS2 dataset compared with their representation in the reference database, Acaulosporaceae, Diversisporaceae, and Gigasporaceae were underrepresented (Fig. S2). The Glomeraceae were the most diverse AM fungal family, with 635 SH in total.

All samples included in the ITS dataset of the GlobalAMFungi database originated from Europe. Among countries, the largest share of samples was gained from Czechia (68%), France (9%), and the United Kingdom (7%). In total, the ITS dataset contained samples from 13 countries. The majority of samples were collected in the grassland biome, and approximately two thirds of the samples originated from plant roots (Fig. 3).

The LSU dataset

We collated 669 samples with data from the focal LSU barcoding region among the 17 selected studies. In total, 1519 747

sequences were assigned to 58 AM fungal taxa (Table 1). The number of sequences assigned to an AM fungal taxon ranged from one to 473 980 per sample, with an average value of 26 202. A total of 24 AM fungal taxa were recorded in ≥ 20 samples, while 21 taxa were found in fewer than 10 samples. Among the phylogenetic lineages (Fig. 2), sequences that were assigned to representative sequences of AM fungal taxa from the family Glomeraceae represented 94% of the whole LSU dataset of the GlobalAMFungi database, followed by Claroideoglomeraceae (4.1%). Other AM fungal families were represented by < 0.5% of the LSU sequences assigned to AM fungal taxa. While Glomeraceae and Claroideoglomeraceae were over-represented than expected by chance in the LSU dataset compared with their representation in the reference database, Acaulosporaceae, Archaeosporaceae, Diversisporaceae, Gigasporaceae, and Paraglomeraceae were underrepresented (Fig. S2). The Glomeraceae were the most diverse AM fungal family with 32 taxa in total.

The majority of samples in the LSU dataset of the GlobalAMFungi database originated from North America and Europe, followed by South America and a few samples from Asia (Fig. 3). Among countries, the largest share of samples was gained from the USA (38%) and Germany (23%). In total, the LSU dataset contained samples from 11 countries. Biomes were predominantly represented by cropland in the LSU dataset, followed by grassland and forest. Sample origin was almost equally split between plant roots and soil (Fig. 3).

Accessing and analysing the dataset through the user interface

The user interface at https://globalamfungi.com enables users to access and analyse the dataset in many ways.

To begin with, the user can apply a 'Taxon search', which offers two different settings: searching for an AM fungal molecular taxon or AM fungal species. The 'Molecular taxon' option limits the search to a particular barcoding region, because different AM fungal barcoding regions use different systems to define and assign AM fungal molecular taxa. While SSU sequences are classified into VT (Öpik *et al.*, 2010), the SH concept was used for ITS2 sequence clusters (Kõljalg *et al.*, 2013) and AM fungal sequences originating from the LSU marker were sorted to molecular taxa that are named according to the GenBank accession numbers of the representative sequences (Delavaux *et al.*, 2021, 2022). By contrast, the 'Species' option enables users to search for any AM fungal species represented in the GlobalAMFungi database across all three barcoding regions. As a result of the 'Taxon search', users get a detailed breakdown of the information about the samples where the taxon was found and about interacting host plant taxa, in case this information is available. The distribution of the taxon (we show that of the AM fungal taxon VTX00215 as an example in Fig. S5a) can be visualized on a map, which also provides approximate information about the relative abundance of the taxon within the AM fungal community. Samples where this taxon was detected can also be sorted based on selected climatic (Fig. S5b) and soil (Fig. S5c) characteristics and compared with the whole GlobalAMFungi dataset to identify the potential ecological preferences of the searched AM fungal taxon. The GlobalAMFungi database also allows postsearch filtering of the data, which, among other things, allows users to ignore samples where the taxon was found as a single sequence (local singleton) or to select a preferred barcoding region(s). The taxon search also enables users to download a FASTA file with all sequence variants assigned to the searched AM fungal molecular taxon. Each sequence variant includes information on sequence ID; sample IDs where the sequence was detected; and the abundance of the sequence in samples in its header.

Second, the user can apply a 'Search by plant', which offers searching for an interacting plant species. This search is limited to SSU dataset only. User can search for 296 plant species. As a result of the 'Search by plant', user gets a detailed breakdown of the information about the samples similar to the 'Taxon search' described above with possibility to download a FASTA file with all sequences variants assigned to AM fungal VT.

Third, our sequence search tools enable users to submit their own sequences and run: a BLAST search for the best hit for up to 100 sequences; or a BLAST search for a single sequence with the option of displaying results for one, 10, 50, 100, or 500 best hits. The resulting pages provide the best hit sequences and the possibility of displaying characteristics of the samples where sequence variants occur, in a similar manner to the results of the 'Taxon search'. Grouped BLAST results provide standard BLAST results with a possibility to filter them according to similarity threshold (Fig. S6). Samples harbouring the closest BLAST hits can be subsequently displayed together with environmental and geographical characteristics similar to the results of a 'Taxon search'.

Lastly, besides searching for individual studies, the user can also select a group of samples directly on the world map using various drawing tools to retrieve the corresponding sample metadata or lists of the AM fungal molecular taxa or species recorded in the selected samples (Fig. S7). In addition to the tools designed for searching within the GlobalAMFungi database, users can also submit their own published data, using the 'Submit your study' option. This takes users to instructions and the web interface for submission.

Discussion

Identifying the environmental drivers of species distribution and predicting distribution shifts represents an important step towards understanding the dynamics of biodiversity under future environmental conditions. Most comprehensive assessments of these topics offer global scale studies (e.g. Kreft & Jetz, 2007; Oliveira *et al.*, 2016; Delgado-Baquerizo *et al.*, 2018; Tedersoo *et al.*, 2022) or databases that compile existing information from various sources (e.g. Jetz *et al.*, 2012; van Kleunen *et al.*, 2019; https://www.gbif.org). Here, we present the newly developed GlobalAMFungi database, which collates and provides information about the distribution of a group of key plant symbionts – AM fungi – that are of profound importance for ecosystem functioning, responding substantially to global change (Baldrian *et al.*, 2022a). The database builds on 100 independent published studies on the composition of AM fungal communities derived using high-throughput sequencing of soil and plant root material

from terrestrial environments, and it is made publicly accessible through a user interface at https://globalamfungi.com.

Recent decades have seen a concerted effort to create databases compiling and sharing information about where and when species have been recorded. One of the largest databases is the Global Biodiversity Information Facility (https://www.gbif.org), which also includes molecular-based data on fungal species distribution from the UNITE database (Nilsson *et al.*, 2019b). Although the UNITE database has revolutionized the field of fungal molecular taxonomy, due to the development and implementation of the SH concept (Kõljalg *et al.*, 2013), it is predominantly built on Sanger-derived nuclear ribosomal ITS sequences mirrored from the International Nucleotide Sequence Databases Collaboration (INSDC), which largely excludes information about fungal species distribution coming from high-throughput sequencing (HTS) methods (but see Tedersoo *et al.*, 2021; Nilsson *et al.*, 2023). The same limitations are also valid for the SILVA database (Quast *et al.*, 2012). While SILVA is a great resource for SSU and LSU rRNA genetic variability over Bacteria, Archaea and Eukaryote domains, it cannot be used as a comprehensive source of information about the distribution of any species/taxa. To overcome this limitation, we recently developed the GlobalFungi database, which builds on collected and validated data published on the composition of fungal communities, using various HTS methods (Větrovský *et al.*, 2020). However, all these databases largely fail in providing comprehensive information about the distribution of AM fungal taxa. One reason for this is that AM fungi are underrepresented in metabarcoding studies focussing on general fungal communities (a particular problem for the GlobalFungi database); a second reason is the existence of multiple barcoding regions for delineating AM fungal taxa (a problem for all databases).

In contrast to studies targeting fungi in general, studies focussing on AM fungal communities most frequently sequence part of the SSU region of rRNA, though to some extent the ITS and the variable part of the LSU region of rRNA are also utilized. This encouraged Öpik *et al.* (2010) to create the MaarjAM database, which gathers molecular-based data on AM fungal species distribution, coming from all three commonly used barcoding regions, for delimitation of AM fungal taxa. Similar to UNITE, the MaarjAM database advanced the field of molecular taxonomy of AM fungi. This was due to the development and implementation of the Virtual Taxon concept for SSU AM data (Öpik *et al.*, 2014), which provides a valuable resource for AM fungal community analyses. However, the MaarjAM database does not fully utilize information about AM fungal species distribution coming from HTS methods. Our newly developed GlobalAMFungi database builds on the molecular taxonomy approaches developed by the teams behind the UNITE and MaarjAM databases, together with a reference dataset of LSU sequences of AM fungal species (Delavaux *et al.*, 2021, 2022) to classify data from 100 studies that used AM fungal-specific primers together with HTS methods to metabarcode AM fungal communities. Similar to the above-listed databases, GlobalAMFungi so far includes only the Glomeromycotinian AM fungal taxa, while the putative AM fungi from the Mucoromycotina subphylum (Orchard *et al.*, 2017; Albornoz *et al.*, 2022) might be included in future updates of the database when the real nature of the symbiosis is described in more detail (e.g. Hoysted *et al.*, 2023). Compared with the above-listed databases, GlobalAMFungi provides orders of magnitude more data on the distribution of AM fungal taxa together with detailed metadata about samples, obtained from original publications or global databases.

The GlobalAMFungi database includes 8486 samples with accurate geographical coordinates from 108 countries. As with other databases that compile fungal occurrence records (Nilsson *et al.*, 2019b; Větrovský *et al.*, 2020; Tedersoo *et al.*, 2021), data in the GlobalAMFungi database are biased towards the traditionally studied regions of Europe and North America. In order to gain a more comprehensive picture about AM fungal diversity and the distribution of AM fungal species on a global scale, the GlobalAMFungi database can be used to identify underrepresented geographical regions that require attention in future sampling campaigns (Averill *et al.*, 2022). Among the most underrepresented regions are tropical areas in South America, Africa, and Southeast Asia. Particularly, striking and concerning is the poor coverage of large areas of temperate grassland in the Middle East and Asia in the GlobalAMFungi database. These ecosystems have given rise to a number of crop plants that underpin the nutrition of a large proportion of the human population and that depend on symbiosis with AM fungi (Wu *et al.*, 2023). Due to agricultural conversion, the area covered by pristine temperate grassland has decreased to < 1% of the original distribution (Scholtz & Twidwell, 2022). Together with these grasslands, we also risk losing information about the AM fungi that coevolved with the ancestors of our crop plants.

To document the potential applicability of the collated dataset to gain new insights into AM fungal ecology, we showed putative preferences of various AM fungal families to the two most commonly sampled substrate types, root and soil. The preference of Glomeraceae, Ambisporaceae, and Pacisporaceae to root over the soil environment corresponds with previously described higher root colonization rate of some species from the Glomeraceae rather than other AM fungal families (Hart & Reader, 2002). On the contrary, Claroideoglomeraceae, Paraglomeraceae, Diversisporaceae, and Archaeosporaceae showed higher relative abundance in soil than plant roots, which indicates their potentially higher investment into extraradical mycelia and spores than intraradical mycelia. Investment into the different structures may determine the relationship of AM fungal species to their environment. While AM fungi investing relatively more into extraradical than intraradical mycelia should be favoured on nutrient-poor sites due to potentially higher ability to benefit their host plants, AM fungi with high intraradical abundance have been proposed to provide the host plant with better protection against root pathogens (Chagnon *et al.*, 2013).

Arbuscular mycorrhizal fungal families also showed different preferences to the studied biomes. The most striking difference was observed between biomes dominated by woody plants on one side and those dominated by herbaceous vegetation and graminoids on the other. Although this observation corresponds to previous evidence identifying large difference in AM fungal

communities between grassland and forest biomes (e.g. Davison *et al.*, 2015), it goes beyond that. Our results indicate that AM fungal families, which favour grassland biome, show also higher relative abundances in croplands compared with woody plant-dominated biomes. This might indicate that plant growth form or lifestyle (perennial vs annual) plays more fundamental role in structuring of AM fungal communities (Davison *et al.*, 2020) than land use intensity. However, the observed patterns should be interpreted with caution, as high variability in biome and substrate type preference may occur on species level.

The GlobalAMFungi database gathers data originating from diverse sources. Due to the absence of a common barcoding region for identification of AM fungal taxa, the GlobalAMFungi database contains sequencing data corresponding to all three (SSU, ITS2, and LSU) commonly used barcoding regions (Öpik *et al.*, 2010). This limits the full potential of the GlobalAMFungi database to only those AM fungal species with known reference sequences for each of the three barcoding regions. Surprisingly, sequences of all three barcoding molecular markers are currently available for only a handful of morphologically defined AM fungal species. Thanks to the independent development of AM fungal molecular taxonomy 'nomenclatures', specific to the SSU (Öpik *et al.*, 2010), ITS (Kõljalg *et al.*, 2013), and LSU (Delavaux *et al.*, 2022) barcoding regions, AM fungal sequences within the GlobalAMFungi database can be sorted into parallel molecular taxonomies. To integrate information about the ecology of AM fungal molecular taxa with information about the taxonomy and phylogeny of AM fungal species, we need a far greater effort in sequencing existing and available AM fungal species.

Outlook

The recent development of long-read sequencing technologies has made it possible to generate high-quality sequences spanning all three commonly used metabarcoding regions of AM fungi (Kolaříková *et al.*, 2021). By sequencing several variable regions together with fairly conserved rRNA genes, these long reads will offer the advantage of robust phylogenetic assignment of short reads obtained from metabarcoding approaches using a single barcoding region. These long amplicon reads also raise the possibility of combining the three currently almost independent datasets within the GlobalAMFungi – based on SSU, ITS, and LSU markers – into a single aggregated database.

The presented version of the GlobalAMFungi database represents a public resource for the scientific community – and a resource we hope will substantially grow over time. We encourage all researchers to upload their own published metabarcoding data on AM fungal communities through the interface on the https://globalamfungi.com Website, following the instructions in the section 'Submit your study'. Besides this, our team will continue to add suitable datasets from published studies. Gathered datasets will be processed, and once per year, a new version of the updated GlobalAMFungi dataset will be released. This routine corresponds to our strategy for maintaining and updating the GlobalFungi database, which grew from > 17 000 samples and > 600 mil. sequences in the first release (Větrovský *et al.*, 2020) to > 57 000 samples and > 3500 mil. sequences in the last release (https://globalfungi.com).

Although the GlobalAMFungi database does not seek to provide resolved taxonomic information for AM fungal sequences, but rather builds on existing molecular taxonomy approaches, the data gathered within the GlobalAMFungi database are freely available to the mycological community and can be easily utilized by AM fungal taxonomists to delimit novel molecular taxa and expand current knowledge about the number of AM fungi. Metadata associated with the sequences can also be examined to assess the ecological niches of existing and novel AM fungal taxa. Identification of distinct ecological niches of closely related novel AM fungal lineages can then provide valuable evidence for AM fungal species delineation. The GlobalAMFungi database is maintained by the same team that created and maintains the GlobalFungi database, which has already been extensively used in various fields, including fungal taxonomy, ecology, and biogeography. Our main goals are to make data about AM fungal species distribution accessible to a range of potential users, to design a user-friendly interface, and to equip the GlobalAMFungi Website with features that allow users to obtain relevant information about the distribution and ecology of AM fungal taxa.

Acknowledgements

We would like to thank all researchers who provided their raw sequencing data and metadata. We are also grateful to Francis Martin, Henrik R. Nilsson, and three anonymous reviewers for their helpful comments that improved the manuscript. This work was supported by the Czech Science Foundation (21-17749S) and by ELIXIR CZ research infrastructure project (MEYS Grant no. LM2023055) through providing access to computing and storage facilities. JD, MM, MO, and MZ were funded by the European Regional Development Fund (Centre of Excellence EcolChange) and the Estonian Research Council (PRG1065 and PRG1789); ZK was supported by the institutional project RVO 67985939.

Competing interests

None declared.

Author contributions

TV, ZK, CL, PB and PK contributed to the study conception and design. TV, ZK, CL, SAH, JD, AF, AG, BJ, MKo, MKr, RJ, LM, TM, MM, AM, SM, IO, MO, MP, SPC, JS, LV, MZ, PB and PK contributed to the data collection. The analyses were performed by TV, ZK, IO and PK. The first draft of the manuscript was written by PK, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

ORCID

Petr Baldrian https://orcid.org/0000-0002-8983-2721
John Davison https://orcid.org/0000-0002-0161-6195

Petr Kohout https://orcid.org/0000-0002-3985-2310
Miroslav Kolařík https://orcid.org/0000-0003-4016-0335
Zuzana Kolaříková https://orcid.org/0000-0002-3693-2657
Manuela Krüger https://orcid.org/0000-0001-9223-3570
Clémentine Lepinay https://orcid.org/0000-0003-1330-1550
Mari Moora https://orcid.org/0000-0002-4819-7506
Maarja Öpik https://orcid.org/0000-0001-8025-7460
Iñaki Odriozola https://orcid.org/0000-0002-5289-7935
Sarah Piché-Choquette https://orcid.org/0000-0002-7447-1177
Tomáš Větrovský https://orcid.org/0000-0002-0831-486X
Lukáš Vlk https://orcid.org/0000-0001-9201-8715
Martin Zobel https://orcid.org/0000-0001-7957-6704

Data availability

All data used in this study are available in the original papers, and compiled dataset is available on https://globalamfungi.com.

References

Albornoz FE, Ryan MH, Bending GD, Hilton S, Dickie IA, Gleeson DB, Standish RJ. 2022. Agricultural land-use favours Mucoromycotinian, but not Glomeromycotinian, arbuscular mycorrhizal fungi across ten biomes. *New Phytologist* **233**: 1369–1382.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Averill C, Anthony MA, Baldrian P, Finkbeiner F, Hoogen J, Kiers T, Kohout P, Hirt E, Smith GR, Crowther TW. 2022. Defending Earth's terrestrial microbiome. *Nature Microbiology* **2022**: 1717–1725.

Baldrian P, Bell-Dereske L, Lepinay C, Větrovský T, Kohout P. 2022a. Fungal communities in soils under global change. *Studies in Mycology* **103**: 1–24.

Baldrian P, Větrovský T, Lepinay C, Kohout P. 2022b. High-throughput sequencing view on the magnitude of global fungal diversity. *Fungal Diversity* **114**: 539–547.

Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, de Wit P, Sánchez-García M, Ebersberger I, de Sousa F *et al.* 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution* **4**: 914–919.

Brundrett MC, Tedersoo L. 2018. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist* **220**: 1108–1115.

Canty A, Ripley BD. 2022. BOOT: bootstrap R (S-Plus) Functions. R package version 1.3-28.1. [WWW document] URL https://cran.r-project.org/web/packages/boot/index.html [accessed 15 May 2023].

Chagnon PL, Bradley RL, Maherali H, Klironomos JN. 2013. A trait-based framework to understand life history of mycorrhizal fungi. *Trends in Plant Science* **18**: 484–491.

Clavel J, Lembrechts J, Alexander J, Haider S, Lenoir J, Milbau A, Nuñez MA, Pauchard A, Nijs I, Verbruggen E. 2021. The role of arbuscular mycorrhizal fungi in nonnative plant invasion along mountain roads. *New Phytologist* **230**: 1156–1168.

Davison J, de Leon DG, Zobel M, Moora M, Bueno CG, Barceló M, Gerz M, León D, Meng Y, Pillar VD *et al.* 2020. Plant functional groups associate with distinct arbuscular mycorrhizal fungal communities. *New Phytologist* **226**: 1117–1128.

Davison J, Moora M, Öpik M, Adholeya A, Ainsaar L, Bâ A, Burla S, Diedhiou AG, Hiiesalu I, Jairus T *et al.* 2015. Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* **349**: 970–973.

Davison J, Moora M, Semchenko M, Adenan SB, Ahmed T, Akhmetzhanova AA, Alatalo JM, al-Quraishy S, Andriyanova E, Anslan S *et al.* 2021. Temperature and pH define the realised niche space of arbuscular mycorrhizal fungi. *New Phytologist* **231**: 763–776.

Davison J, Öpik M, Zobel M, Vasar M, Metsis M, Moora M. 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.

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.

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 Phytologist* **229**: 3048–3052.

Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in soil. *Science* **359**: 320–325.

Fan K, Delgado-Baquerizo M, Yguan Z, Chu H. 2020. Crop production correlates with soil multitrophic communities at the large spatial scale. *Soil Biology and Biochemistry* **151**: 108047.

Hart M, Reader RJ. 2002. Taxonomic basis for variation in the colonization strategy of arbuscular mycorrhizal fungi. *New Phytologist* **153**: 335–344.

Hengl T, De Jesus JM, Heuvelink GBM, Gonzalez MR, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B *et al.* 2017. SoilGrids250m: global gridded soil information based on machine learning. *PLoS ONE* **12**: e0169748.

House GL, Bever JD. 2020. Biochar soil amendments in prairie restorations do not interfere with benefits from inoculation with native arbuscular mycorrhizal fungi. *Restoration Ecology* **28**: 785–795.

Hoysted GA, Field KJ, Sinanaj B, Bell CA, Bidartondo MI, Pressel S. 2023. Direct nitrogen, phosphorus and carbon exchanges between Mucoromycotina 'fine root endophyte' fungi and a flowering plant in novel monoxenic cultures. *New Phytologist* **238**: 70–79.

Jetz W, Mcpherson JM, Guralnick RP. 2012. Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution* **27**: 151–159.

Jiang S, Liu Y, Luo J, Qin M, Johnson NC, Öpik M, Vasar M, Chai Y, Zhou X, Mao L *et al.* 2018. Dynamics of arbuscular mycorrhizal fungal community structure and functioning along a nitrogen enrichment gradient in an alpine meadow ecosystem. *New Phytologist* **220**: 1222–1235.

Kakouridis A, Hagen JA, Kan MP, Mambelli S, Feldman LJ, Herman DJ, Weber PK, Pett-Ridge J, Firestone MK. 2022. Routes to roots: direct evidence of water transport by arbuscular mycorrhizal fungi to host plants. *New Phytologist* **236**: 210–221.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* **4**: 1–20.

Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A *et al.* 2011. Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* **333**: 880–882.

van Kleunen M, Pyšek P, Dawson W, Essl F, Kreft H, Pergl J, Weigelt P, Stein A, Dullinger S, König C *et al.* 2019. The Global Naturalized Alien Flora (GloNAF) database. *Ecology* **100**: e02542.

Koch AM, Antunes PM, Maherali H, Hart MM, Klironomos JN. 2017. Evolutionary asymmetry in the arbuscular mycorrhizal symbiosis: conservatism in fungal morphology does not predict host plant growth. *New Phytologist* **214**: 1330–1337.

Kohout P, Doubková P, Bahram M, Suda J, Tedersoo L, Voříšková J, Sudová R. 2015. Niche partitioning in arbuscular mycorrhizal communities in temperate grasslands: a lesson from adjacent serpentine and nonserpentine habitats. *Molecular Ecology* **24**: 1831–1843.

Kohout P, Sudová R, Janoušková M, Čtvrtlíková M, Hejda M, Pánková H, Slavíková R, Štajerová K, Vosátka M, Sýkorová Z. 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.

Kolaříková Z, Slavíková R, Krüger C, Krüger M, Kohout P. 2021. PacBio sequencing of Glomeromycota rDNA: a novel amplicon covering all widely

used ribosomal barcoding regions and its applicability in taxonomy and ecology of arbuscular mycorrhizal fungi. *New Phytologist* **231**: 490–499.

Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM *et al.* 2013. Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology* **22**: 5271–5277.

Kreft H, Jetz W. 2007. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences, USA* **104**: 5925–5930.

Krüger M, Teste FP, Laliberté E, Lambers H, Coghlan M, Zemunik G, Bunce M. 2015. The rise and fall of arbuscular mycorrhizal fungal diversity during ecosystem retrogression. *Molecular Ecology* **24**: 4912–4930.

Lahti L, Shetty S. 2018. Introduction to the microbiome R package. [WWW document] URL https://microbiome.github.io/tutorials/ [accessed 13 December 2021].

Lekberg Y, Gibbons SM, Rosendahl S, Ramsey PW. 2013. Severe plant invasions can increase mycorrhizal fungal abundance and diversity. *The ISME Journal* **7**: 1424–1433.

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 Phytologist* **220**: 971–976.

Maestre FT, Delgado-Baquerizo M, Jeffries TC, Eldridge DJ, Ochoa V, Gozalo B, Quero JL, García-Gómez M, Gallardo A, Ulrich W *et al.* 2015. Increasing aridity reduces soil microbial diversity and abundance in global drylands. *Proceedings of the National Academy of Sciences, USA* **112**: 15684–15689.

Moulíková Š, Kolařík M, Lorch JM, Kolarczyková D, Hubka V, Čmoková A. 2023. Wild rodents harbour high diversity of Arthroderma. *Persoonia – Molecular Phylogeny and Evolution of Fungi* **50**: 27–47.

Mozzachiodi S, Bai FY, Baldrian P, Bell G, Boundy-Mills K, Buzzini P, Čadež N, Cubillos FA, Dashko S, Dimitrov R *et al.* 2022. Yeasts from temperate forests. *Yeast* **39**: 4–24.

Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. 2019a. Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews. Microbiology* **17**: 95–109.

Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L *et al.* 2019b. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research* **47**: D259–D264.

Nilsson RH, Ryberg M, Wurzbacher C, Tedersoo L, Anslan S, Põlme S, Spirin V, Mikryukov V, Svantesson S, Hartmann M *et al.* 2023. What mycologists should talk about when they are talking about the International Code of Nomenclature for algae, fungi, and plants. *bioRxiv*. doi: 10.1101/2023.02.25.529721.

Ohsowski BM, Zaitsoff PD, Öpik M, Hart MM. 2014. Where the wild things are: looking for uncultured Glomeromycota. *New Phytologist* **204**: 171–179.

Oliveira BF, Machac A, Costa GC, Brooks TM, Davidson AD, Rondinini C, Graham CH. 2016. Species and functional diversity accumulate differently in mammals. *Global Ecology and Biogeography* **25**: 1119–1130.

Öpik M, Davison J. 2016. Uniting species- and community-oriented approaches to understand arbuscular mycorrhizal fungal diversity. *Fungal Ecology* **24**: 106–113.

Öpik M, Davison J, Moora M, Zobel M. 2014. DNA-based detection and identification of Glomeromycota: the virtual taxonomy of environmental sequences. *Botany* **92**: 135–147.

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

Öpik M, Zobel M, Cantero JJ, Davison J, Facelli JM, Hiiesalu I, Jairus T, Kalwij JM, Koorem K, Leal ME *et al.* 2013. Global sampling of plant roots expands the described molecular diversity of arbuscular mycorrhizal fungi. *Mycorrhiza* **23**: 411–430.

Orchard S, Hilton S, Bending GR, Dickie IA, Standish RJ, Gleeson DB, Jeffery RP, Powell JR, Walker C, Bass D *et al.* 2017. Fine endophytes (*Glomus tenue*) are related to Mucoromycotina, not Glomeromycota. *New Phytologist* **213**: 481–486.

Qiu Q, Bender SF, Mgelwa AS, Hu Y. 2022. Arbuscular mycorrhizal fungi mitigate soil nitrogen and phosphorus losses: a meta-analysis. *Science of the Total Environment* **807**: 150857.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**: D590–D596.

Réblová M, Kolařík M, Nekvindová J, Réblová K, Sklenář F, Miller AN, Hernández-Restrepo M. 2021. Phylogenetic reassessment, taxonomy, and biogeography of Codinaea and similar fungi. *Journal of Fungi* **20**: 1097.

Rodriguez-Ramos JC, Cale JA, Cahill JF, Simard SW, Karst J, Erbilgin N. 2021. Changes in soil fungal community composition depend on functional group and forest disturbance type. *New Phytologist* **229**: 1105–1117.

Sato K, Suyama Y, Saito M, Sugawara K. 2005. A new primer for discrimination of arbuscular mycorrhizal fungi with polymerase chain reaction-denature gradient gel electrophoresis. *Grassland Science* **51**: 179–181.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences, USA* **109**: 6241–6246.

Scholtz R, Twidwell D. 2022. The last continuous grasslands on Earth: identification and conservation importance. *Conservation Science and Practice* **4**: e626.

Smith S, Read D. 2008. *Mycorrhizal symbiosis*. Cambridge, UK: Academic Press.

Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A *et al.* 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**: 1028–1046.

Stürmer SL, Bever JD, Morton JB. 2018. Biogeography of arbuscular mycorrhizal fungi (Glomeromycota): a phylogenetic perspective on species distribution patterns. *Mycorrhiza* **28**: 587–603.

Stürmer SL, Bever JD, Schultz PA, Bentivenga SP. 2021. Celebrating INVAM: 35 years of the largest living culture collection of arbuscular mycorrhizal fungi. *Mycorrhiza* **31**: 117–126.

Sugiyama Y, Sato H. 2021. The limited establishment of native ectomycorrhizal fungi in exotic *Eucalyptus* spp. stands in Japan. *Frontiers in Microbiology* **12**: 376.

Tedersoo L, Bahram M, Polme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A *et al.* 2014. Global diversity and geography of soil fungi. *Science* **346**: 1256688.

Tedersoo L, Mikryukov V, Anslan S, Bahram M, Khalid AN, Corrales A, Agan A, Vasco-Palacios AM, Saitta A, Antonelli A *et al.* 2021. The Global Soil Mycobiome consortium dataset for boosting fungal diversity research. *Fungal Diversity* **111**: 573–588.

Tedersoo L, Mikryukov V, Zizka A, Bahram M, Hagh-Doust N, Anslan S, Prylutskyi O, Delgado-Baquerizo M, Maestre FT, Pärn J *et al.* 2022. Global patterns in endemicity and vulnerability of soil fungi. *Global Change Biology* **28**: 6696–6710.

Van Geel M, Busschaert P, Honnay O, Lievens B. 2014. Evaluation of six primer pairs targeting the nuclear rRNA operon for characterization of arbuscular mycorrhizal fungal (AMF) communities using 454 pyrosequencing. *Journal of Microbiological Methods* **106**: 93–100.

Vasar M, Andreson R, Davison J, Jairus T, Moora M, Remm M, Young JPW, Zobel M, Öpik M. 2017. Increased sequencing depth does not increase captured diversity of arbuscular mycorrhizal fungi. *Mycorrhiza* **27**: 761–773.

Vasar M, Davison J, Sepp SK, Öpik M, Moora M, Koorem K, Meng Y, Oja J, Akhmetzhanova AA, al-Quraishy S *et al.* 2021. Arbuscular mycorrhizal fungal communities in the soils of desert habitats. *Microorganisms* **9**: 1–14.

Větrovský T, Baldrian P, Morais D. 2018. SEED 2: a user-friendly platform for amplicon high-throughput sequencing data analyses. *Bioinformatics* **34**: 2292–2294.

Větrovský T, Kohout P, Kopecký M, Machac A, Man M, Bahnmann BD, Brabcová V, Choi J, Meszárošová L, Human ZR *et al.* 2019. A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nature Communications* **10**: 5142.

Větrovský T, Morais D, Kohout P, Lepinay C, Algora C, Awokunle Hollá S, Bahnmann BD, Bílohnědá K, Brabcová V, D'Alò F *et al.* 2020. GlobalFungi,

a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. *Scientific Data* 7: 1–14.

Vlk L, Tedersoo L, Antl T, Větrovský T, Abarenkov K, Pergl J, Albrechtová J, Vosátka M, Baldrian P, Pyšek P *et al.* 2020a. Alien ectomycorrhizal plants differ in their ability to interact with co-introduced and native ectomycorrhizal fungi in novel sites. *The ISME Journal* **14**: 2336–2346.

Vlk L, Tedersoo L, Antl T, Větrovský T, Abarenkov K, Pergl J, Albrechtová J, Vosátka M, Baldrian P, Pyšek P *et al.* 2020b. Early successional ectomycorrhizal fungi are more likely to naturalize outside their native range than other ectomycorrhizal fungi. *New Phytologist* **227**: 1289–1293.

Wijayawardene NN, Hyde KD, Al-Ani LKT, Tedersoo L, Haelewaters D, Rajeshkumar KC, Zhao RL, Aptroot A, Leontyev DV, Saxena RK *et al.* 2020. Outline of fungi and fungus-like taxa. *Mycosphere* **11**: 1060–1456.

Wu H, Yang J, Fu W, Rillig MC, Cao Z, Zhao A, Hao Z, Zhang X, Chen B, Han X. 2023. Identifying thresholds of nitrogen enrichment for substantial shifts in arbuscular mycorrhizal fungal community metrics in a temperate grassland of northern China. *New Phytologist* **237**: 279–294.

Xiang D, Verbruggen E, Hu Y, Veresoglou SD, Rillig MC, Zhou W, Xu T, Li H, Hao Z, Chen Y *et al.* 2014. Land use influences arbuscular mycorrhizal fungal communities in the farming–pastoral ecotone of northern China. *New Phytologist* **204**: 968–978.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Geographic distribution of the samples within the GlobalAMFungi database.

Fig. S2 Arbuscular mycorrhizal fungal families over-represented (bars above and not overlapping with the dashed line) or underrepresented (bars below and not overlapping with the dashed line) in the three datasets (SSU, ITS2, and LSU) within the GlobalAMFungi database.

Fig. S3 Differences of relative abundances of different AM fungal families between biomes, based on the SSU dataset.

Fig. S4 Differences of relative abundances of different AM fungal families between substrate types, based on the SSU dataset.

Fig. S5 Results of the 'Taxon search' tool in the GlobalAMFungi database for the example taxon VTX00215.

Fig. S6 Screenshot of the 'Sequence search' result in the GlobalAMFungi database.

Fig. S7 Results of the 'Geosearch' function.

Table S1 List of all primers, their sequences, and references from the collated papers.

Table S2 List of metadata contained in the GlobalAMFungi database.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.