Coming up short: Identifying substrate and geographic biases in fungal sequence databases

Maryia Khomich[1,2*], Filipa Cox[3,4], Carrie J. Andrew[5,2], Tom Andersen[1], Håvard Kauserud[2], Marie L. Davey[2,6]

[1]Section for Aquatic Biology and Toxicology, Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway

[2]Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway

[3]School of Earth and Environmental Sciences, University of Manchester, Manchester M13 9PL, United Kingdom

[4]British Antarctic Survey, Natural Environment Research Council, Cambridge CB3 0ET, United Kingdom

[5]Swiss Federal Institute for Forest, Snow, and Landscape Research WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland

[6]Norwegian Institute of Bioeconomy Research, Department of Soil Quality and Climate Change, Høgskoleveien 7, 1430 Ås, Norway

*Corresponding author: Maryia Khomich, Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway

Tel.: +47-22845979, Fax: +47-22854726

E-mail address: maryia.khomich@ibv.uio.no, marykhomich@gmail.com

Abstract

Insufficient reference database coverage is a widely recognized limitation of molecular ecology approaches which are reliant on database matches for assignment of function or identity. Here, we use data from 65 amplicon high-throughput sequencing (HTS) datasets targeting the internal transcribed spacer (ITS) region of fungal rDNA to identify substrates and geographic areas whose underrepresentation in the available reference databases could have meaningful impact on our ability to draw ecological conclusions. A total of 14 different substrates were investigated. Database representation was particularly poor for the fungal communities found in aquatic (freshwater and marine) and soil ecosystems. Aquatic ecosystems are identified as priority targets for the recovery of novel fungal lineages. A subset of the data representing soil samples with global distribution were used to identify geographic locations and terrestrial biomes with poor database representation. Database coverage was especially poor in tropical, subtropical, and Antarctic latitudes, and the Amazon, Southeast Asia, Australasia, and the Indian subcontinent are identified as priority areas for improving database coverage in fungi.

Keywords: fungi, mycobiome, ITS region, GenBank, UNITE, RDP Bayesian classifier, diversity, metabarcoding, substrate

Introduction

Fungi encompass one of the most functionally and ecologically diverse kingdoms of eukaryotes, maintaining ecosystem functioning on a global scale and playing fundamental roles as decomposers, mutualists and pathogens of animals and plants (Peay *et al.*, 2016). Estimates of global fungal diversity range from 0.6 to 5.1 million species of fungi (Hawksworth, 2001, Bass & Richards, 2011, Blackwell, 2011, Hawksworth, 2012). However, to date, only a tiny fraction of them (ca. 140 000 species) have been classified, although some 1200 new fungal species are described each year (Kirk *et al.*, 2008, Hibbett *et al.*, 2011).

The advent of massively parallel high-throughput sequencing (HTS) has enabled the exploration of fungal diversity on a previously impossible scale (Hibbett *et al.*, 2009). As a result, fungal barcoding of environmental samples is increasingly driving the exploration of the processes structuring fungal diversity, the identification of ecosystem functions linked to fungal diversity and the discovery of novel fungal biodiversity, especially for understudied geographic regions and substrates (Schoch *et al.*, 2012, Öpik *et al.*, 2016). Fungal barcoding approaches largely focus on the internal transcribed spacer (ITS) region, which is the standard barcode for Fungi (Schoch *et al.*, 2012). The establishment of large-scale public reference ITS databases is therefore crucial to allow reliable sequence-based identification of fungal species in HTS approaches (Coissac *et al.*, 2016).

Database-dependent HTS approaches suffer from several biases and limitations directly related to the quality and breadth of the databases. For example, there are only a relatively small fraction of reference database sequences for which a specimen or culture is readily available (Bridge *et al.*, 2003) and consequently large proportions of environmental sequences typically are not represented in the sequence databases. In the case of Fungi, the three public repositories

in the International Nucleotide Sequence Database Collaboration (INSDC), namely the DNA Data Bank of Japan (DDBJ), European Nucleotide Archive (ENA) and GenBank, have become a default resource of taxonomic annotation for newly generated environmental sequences (Karsch-Mizrachi *et al.*, 2018). However, it has been reported that 10-21% of fungal sequences deposited in INSD can be either chimeric, of poor quality or contain incorrect and insufficient taxonomic information (Bridge *et al.*, 2003, Nilsson *et al.*, 2006). To improve the annotation of fungal ITS sequences from NCBI databases, the ITS RefSeq Targeted Loci project has been initiated to develop a separate, curated database representing sequences from type material and stored in public archives (Schoch *et al.*, 2014, Robbertse *et al.*, 2017). By contrast, UNITE (unite.ut.ee) provides highly filtered, curated ITS reference sequences for molecular identification of fungi (Kõljalg *et al.*, 2013). The geographic representation in both databases is strongly skewed towards Europe, North America, China, and Japan (Ryberg *et al.*, 2009, Kõljalg *et al.*, 2013). As a result, satisfactory taxonomic assignment remains problematic in the kingdom Fungi due to the lack of reliable and correctly annotated reference sequences, and coverage related biases.

Here, we assess the impact of unbalanced database representation by geographic locale and substrate on our ability to discern and identify the components of fungal communities using data from 65 amplicon HTS datasets targeting the ITS region of rDNA. We attempt to identify both substrates and geographic regions in which underrepresentation in the available fungal ITS databases could have meaningful impact on our interpretation of HTS amplicon sequencing data.

Materials and Methods

The data analysed represent 65 next generation ITS amplicon sequencing datasets from 14 different substrates, including terrestrial, aquatic, and marine environments, as well as plant and

animal hosts (Table S1). Data were gleaned from published materials, through personal communication with the authors, or from public data archives (e.g. ENA or NCBI Sequence Read Archive (SRA)). Among these datasets, 30 were derived from soil substrates representing 625 sites from 14 biomes with global distribution across all continents (Table S1, Table S2). Sites were assigned to biomes following the classification of the World Wildlife Foundation (http://worldwildlife.org) with the following modifications: (i) temperate deciduous forests in the Northern and Southern hemispheres were treated separately; (ii) montane forests were separated from lowland forests in the tropics; (iii) grasslands and shrublands were considered as a single unit globally, and (iv) vegetated subantarctic sites were differentiated from unvegetated maritime Antarctic sites. For all datasets, sequences were error-corrected and quality-filtered prior to clustering into operational taxonomic units (OTUs) at a 97% similarity threshold (Table S1). Although intraspecific ITS variability ranges from zero to 24.2% (Nilsson *et al.*, 2008), the 97% threshold is widely used to delineate fungal OTUs at approximately species level (Hughes *et al.*, 2009, Ryberg, 2015), and the use of a single threshold across all datasets here allows for comparison across studies and geographic areas (Yahr *et al.*, 2016). Global singletons were considered probable sequencing errors and removed (Quince *et al.*, 2009, Kunin *et al.*, 2010, Tedersoo *et al.*, 2010), as were chimeric sequences. We limited analyses to only those OTUs with representative sequences > 99 nt in length, as suggested by Tedersoo *et al.* (2014). BLAST searches of the representative sequences of each OTU were made against the reference databases NCBI-nr/nt (v.2.2.29) (hereafter referred to as NCBI) and UNITE v. 7 (unite.ut.ee). An alternative taxonomic assignment method, the RDP Naïve Bayesian rRNA Classifier (v. 2.11), was used to query the representative sequences against a UNITE+INSD-based database, the Warcup Fungal ITS training set 2 (Deshpande *et al.*, 2016). The method employs multiple

hierarchy models for several gene regions, including ITS, to bootstrap 8 nt *k*-mers of the query sequence against the reference dataset and calculate an assignment score for each taxonomic rank (Deshpande *et al.*, 2016). OTUs were considered to be non-target and discarded if they either: (i) were identified as a non-fungal organism by any database or (ii) their best BLAST match to a fungal reference sequence had a query coverage of < 70%. In total 196 790 OTUs were analysed (Fig. S1, Appendix 1).

The relative representation of environmental sequences in the NCBI, UNITE, and RDP Warcup databases was assessed across substrates, biomes, and geographic locations. OTUs were considered to be represented (i.e. to have a match) in the NCBI or UNITE databases if the representative sequence had a BLAST match of > 97% identity to a sequence in the reference database. OTUs classified with > 80% confidence at a given taxonomic level using the RDP classifier were considered to be successfully taxonomically assigned. We calculated the proportion of OTUs from each substrate that were represented in the NCBI and UNITE databases, as well as the proportion of OTUs that could be successfully assigned at the phylum and genus levels using the RDP classifier. Using only the soil-inhabiting OTUs from the dataset, we also calculated the proportion of OTUs from distinct biomes that were represented in the databases and could be successfully assigned to phylum and genus. Patterns in the proportion of soil-inhabiting OTUs represented in the databases and successfully assigned at the genus level relative to latitude and longitude were investigated by fitting up to third order polynomial functions and selecting best fit models on the basis of AICc values. To account for unequal sampling intensity, observations were weighted by sampling frequency within 0.1 degree latitudinal and longitudinal ranges. To further assess geographic patterns in successful assignment of soil inhabiting OTUs, inverse distance weighting (IDW) spatial interpolation of (i)

representation in the NCBI database, (ii) representation in the UNITE database and (iii) successful assignment at the genus level were used to estimate global database coverage. Confidence intervals for each interpolation were calculated using a jackknife estimator with 100 permutations.

Results and Discussion

On average, more OTUs were represented in NCBI than in UNITE, which is unsurprising considering the substantial size difference in the two databases. However, it must be noted that representation only denotes the existence of a similar, previously deposited sequence in the database and does not guarantee successful assignment at a given taxonomic level (Fig. 1, Fig. S2-S3, Table S1). A substantial proportion of the OTUs of most datasets could be assigned at the phylum level (mean=0.80, range=0.43-0.95), but assignment success decreased substantially at lower taxonomic levels (i.e. genus: mean=0.42, range=0.24-0.57) (Fig. S4). Across all datasets, the most OTU-rich, and therefore presumably most speciose fungal phyla were Ascomycota (51%) and Basidiomycota (39%). These lineages were also more successfully assigned using the RDP's ITS fungal training set 2, at both the phylum (Ascomycota: mean=0.88, range=0.66–0.97; Basidiomycota: mean=0.78, range=0.38–0.95) and genus levels (Ascomycota: mean=0.40, range=0.19–0.56; Basidiomycota: mean=0.49, range=0.30–0.64) (Fig. S5-S6). Better representation and assignment of Dikarya compared to basal fungal lineages can likely be attributed to both primer bias in the commonly used ITS barcoding primer pairs (Bokulich & Mills, 2013, Tedersoo & Lindahl, 2016) and the Dikarya-biased taxonomic composition of the reference databases.

The proportions of OTUs represented in the databases and that could be successfully

taxonomically assigned varied across all substrates, but were comparatively low in aquatic environments and soil. Only 31-46% of OTUs in these substrates were represented in the NCBI/UNITE databases, and in marine and freshwater substrates in particular, only approximately half of the fungal OTUs could be successfully assigned at the phylum level (Fig. 1, Fig. S4). The combined poor representation and lower degree of successful phylum level assignments may suggest that marine and aquatic environments host a higher proportion of novel, unclassified, and yet undescribed fungal lineages. In particular, it is thought that aquatic habitats represent a larger fraction of unknown fungal diversity than previously acknowledged (Richards *et al.*, 2012). Despite overall low representation in the reference databases, soil had higher success rates of taxonomic assignment at the phylum level (66%; Fig. 1) than aquatic and marine environments. This likely reflects both a highly diverse fungal community from known lineages in soil (de Boer *et al.*, 2005), as well as database related biases due to the large number of 'unnamed environmental sequences' deposited in both NCBI and UNITE that preclude taxonomic assignment (Hibbett *et al.*, 2011). Improved database coverage in fungi will clearly require both novel lineage characterization and attempts to link environmental sequences with identified organisms.

Although the exploration of fungal diversity across habitats worldwide has been greatly facilitated by the development of HTS approaches, the vast majority of fungal species and their distribution for most geographic regions remains unknown (Schoch *et al.*, 2012). With the rise of molecular and HTS tools for biodiversity exploration, database breadth and quality have become integral in ensuring successful and meaningful data interpretation in these studies. In order to investigate database representation from a geographic perspective, we analysed a subset of the datasets representing soils from 625 sampling sites encompassing 14 terrestrial biomes

worldwide (Fig. 2 and 3, Appendix 2)**.** Database representation (i.e. the proportion of OTUs with a high quality match to a pre-existing reference sequence) varied between biomes. This suggests that: (1) fungal diversity has not been consistently explored, inventoried, and databased across all biomes and (2) there accordingly will be biome-specific biases in our ability to extract reliable database related information about fungal communities including taxonomy, guild, and trait information. Boreal, temperate coniferous, temperate deciduous and tropical montane forests are the biomes with best database representation, while dunes, mangroves, savannas and the subantarctic are among the terrestrial biomes with the poorest database representation (Fig. 2). The proportion of OTUs successfully assigned to phylum was consistent across both latitude and biomes, with the exception of the moist tropical forests. This pattern would seem to suggest that those geographic areas underrepresented in the databases harbour additional diversity among known lineages, rather than a high proportion of novel lineages forming deep branches in the tree of Kingdom Fungi (Fig. 2). The best model explaining database representation in both NCBI and UNITE, and genus-level RDP assignment was in all cases a third order polynomial fit of latitude (R^2= 0.60, R^2= 0.34, R^2= 0.20 respectively; Fig. 4) which exhibited a clear drop in database representation towards the equator and Antarctic. No models found significant fits for longitude as an explanatory variable (data not shown). Database representation was highest in temperate, northern hemisphere latitudes while the tropics and subtropics were most poorly represented (Fig. 3, Fig. S7-S9). This trend likely reflects a combined effect of the comparatively high richness of soil fungi in the tropics in combination with lower sampling effort compared to North America and Europe (Ryberg *et al.*, 2009). IDW identified the Amazon, Australasia, Southeast Asia and the Indian subcontinent as being particularly underrepresented in existing databases, suggesting that these geographic areas should be prioritized to improve coverage in existing

databases.

In conclusion, with the increasing use of database dependent HTS approaches to address questions in fungal biodiversity and ecology, reference database quality is becoming an increasingly pressing concern. Public databases are dynamic and their quality has consistently improved with time and concerted effort by both users and developers (Nilsson *et al.*, 2014, O'Leary *et al.*, 2015). We have identified both priority substrates and geographic regions to which efforts may be focused to most efficiently advance improvement in database coverage. The combined low database representation and high proportions of OTUs that cannot be classified at high taxonomic levels observed in freshwater and marine substrates suggests they are the most likely candidates for recovery of novel lineages representing deep branches within Kingdom Fungi. Northern and temperate biomes are best represented in the databases, and tropical regions, including the Amazon and Southeast Asia, are identified as priority areas for improving global database coverage.

Acknowledgements

This study has been supported financially by the Department of Biosciences, University of Oslo and by the Research Council of Norway (grant 'COMSAT' 196336 to Tom Andersen). We thank all authors who were involved in data generation used in the meta-analysis. All data for this manuscript is properly cited and referred to in the Table S1.

References

Bass D & Richards TA (2011) Three reasons to re-evaluate fungal diversity 'on Earth and in the ocean'. *Fungal Biology Reviews* **25**: 159-164.

Blackwell M (2011) The Fungi: 1, 2, 3… 5.1 million species? *American Journal of Botany* **98**: 426-438.

Bokulich NA & Mills DA (2013) Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Applied and environmental microbiology* **79**: 2519-2526.

Bridge PD, Roberts PJ, Spooner BM & Panchal G (2003) On the unreliability of published DNA sequences. *New Phytologist* **160**: 43-48.

Coissac E, Hollingsworth PM, Lavergne S & Taberlet P (2016) From barcodes to genomes: extending the concept of DNA barcoding. *Molecular ecology*.

de Boer W, Folman LB, Summerbell RC & Boddy L (2005) Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS microbiology reviews* **29**: 795-811.

Deshpande V, Wang Q, Greenfield P, Charleston M, Porras-Alfaro A, Kuske CR, Cole JR, Midgley DJ & Tran-Dinh N (2016) Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia* **108**: 1-5.

Hawksworth D (2012) Global species numbers of fungi: are tropical studies and molecular approaches contributing to a more robust estimate? *Biodiversity and Conservation* **21**: 2425-2433.

Hawksworth DL (2001) The magnitude of fungal diversity: the 1· 5 million species estimate revisited. *Mycological research* **105**: 1422-1432.

Hibbett DS, Ohman A & Kirk PM (2009) Fungal ecology catches fire. *New Phytologist* **184**: 279-282.

Hibbett DS, Ohman A, Glotzer D, Nuhn M, Kirk P & Nilsson RH (2011) Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. *Fungal Biology Reviews* **25**: 38-47.

Hughes KW, Petersen RH & Lickey EB (2009) Using heterozygosity to estimate a percentage DNA sequence similarity for environmental species' delimitation across basidiomycete fungi. *New Phytologist* **182**: 795-798.

Karsch-Mizrachi I, Takagi T & Cochrane G (2018) The international nucleotide sequence database collaboration. *Nucleic Acids Res* **46**: D48-D51.

Kirk P, Cannon P, Minter D & Stalpers J (2008) *Dictionary of the Fungi, 10th Edition.* CABI, Wallingford, UK.

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

Kunin V, Engelbrektson A, Ochman H & Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**: 118-123.

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N & Larsson K-H (2008) Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary Bioinformatics Online* **4**: 193-201.

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H & Kõljalg U (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS One* **1**: e59.

Nilsson RH, Hyde KD, Pawłowska J, Ryberg M, Tedersoo L, Aas AB, Alias SA, Alves A, Anderson CL & Antonelli A (2014) Improving ITS sequence data for identification of plant pathogenic fungi. *Fungal Diversity* **67**: 11-19.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B & Ako-Adjei D (2015) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research* **44**: D733-D745.

Peay KG, Kennedy PG & Talbot JM (2016) Dimensions of biodiversity in the Earth mycobiome. *Nature Reviews Microbiology* **14**: 434-447.

Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF & Sloan WT (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods* **6**: 639-641.

Richards TA, Jones MD, Leonard G & Bass D (2012) Marine fungi: their ecology and molecular diversity. *Annual Review of Marine Science* **4**: 495-522.

Robbertse B, Strope PK, Chaverri P, Gazis R, Ciufo S, Domrachev M & Schoch CL (2017) Improving taxonomic accuracy for fungi in public sequence databases: applying 'one name one species' in well-defined genera with Trichoderma/Hypocrea as a test case. *Database* **2017**.

Ryberg M (2015) Molecular operational taxonomic units as approximations of species in the light of evolutionary models and empirical data from Fungi. *Molecular ecology* **24**: 5770-5777.

Ryberg M, Kristiansson E, Sjökvist E & Nilsson RH (2009) An outlook on the fungal internal transcribed spacer sequences in GenBank and the introduction of a web-based tool for the exploration of fungal diversity. *New Phytologist* **181**: 471-477.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K & Crous PW (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* **109**: 6241-6246.

Schoch CL, Robbertse B, Robert V, Vu D, Cardinali G, Irinyi L, Meyer W, Nilsson RH, Hughes K & Miller AN (2014) Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi. *Database* **2014**: bau061.

Tedersoo L & Lindahl B (2016) Fungal identification biases in microbiome projects. *Environmental Microbiology Reports* **8**: 774-779.

Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G & Kõljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist* **188**: 291-301.

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

Yahr R, Schoch CL & Dentinger BT (2016) Scaling up discovery of hidden diversity in fungi: impacts of barcoding approaches. *Phil Trans R Soc B* **371**: 20150336.

Öpik M, Davison J, Moora M, Pärtel M & Zobel M (2016) Response to Comment on "Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism". *Science* **351**: 826-826.

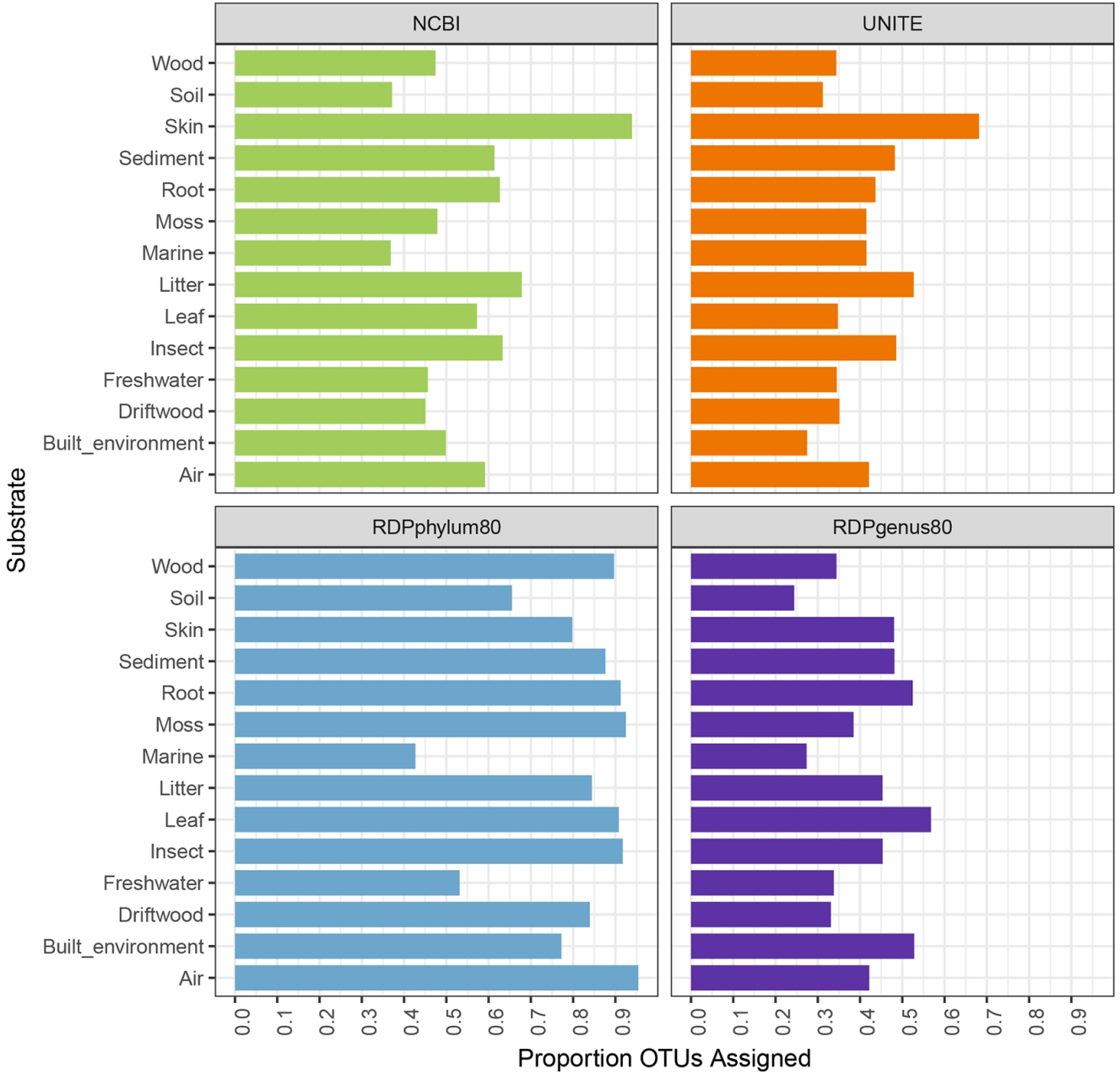
Figure legends

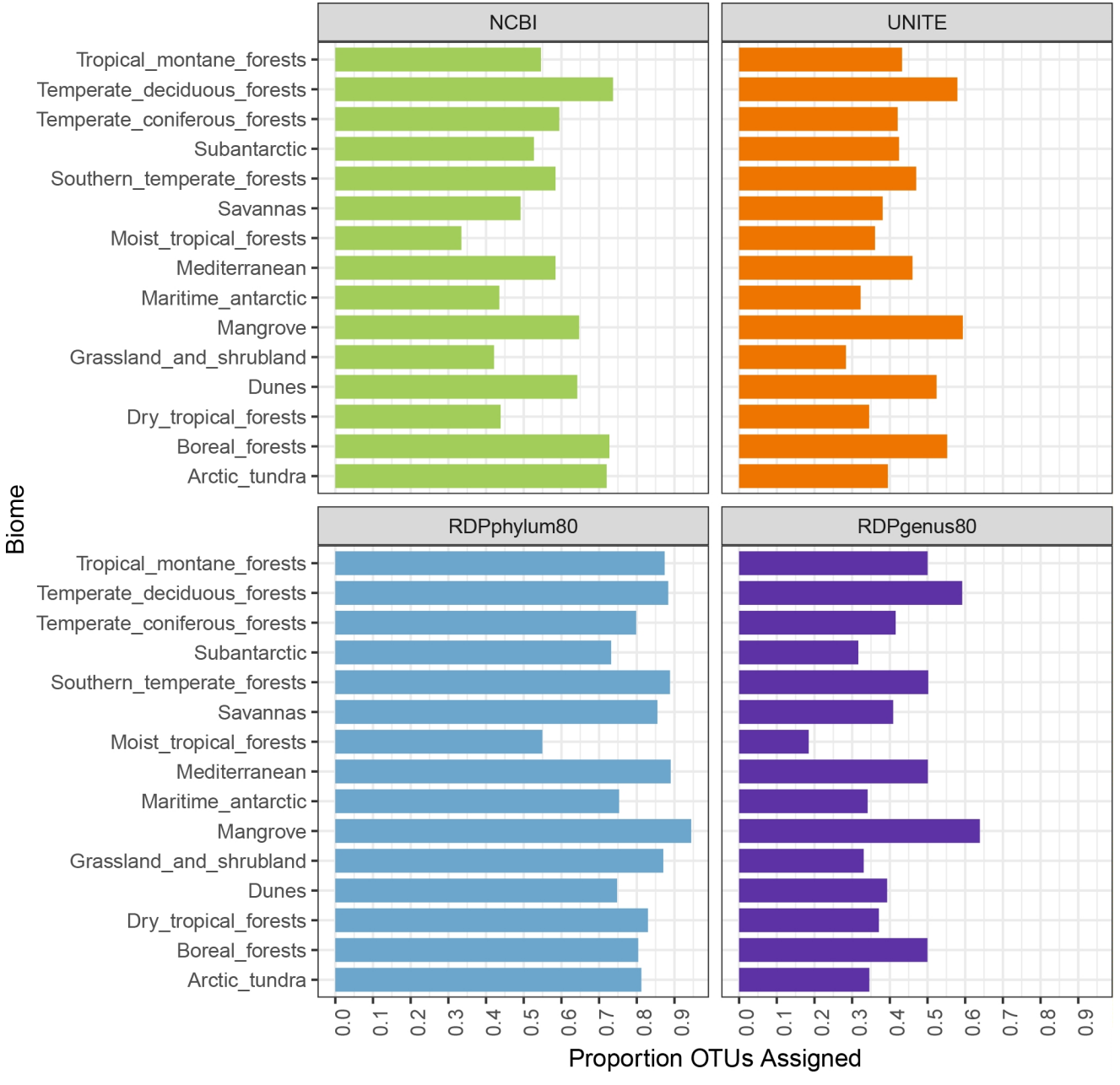
Figure 1. A barplot representing proportion of assigned fungal OTUs across 14 different substrates in NCBI (green), UNITE (orange) and RDP (phylum level: blue; genus level: purple) databases.

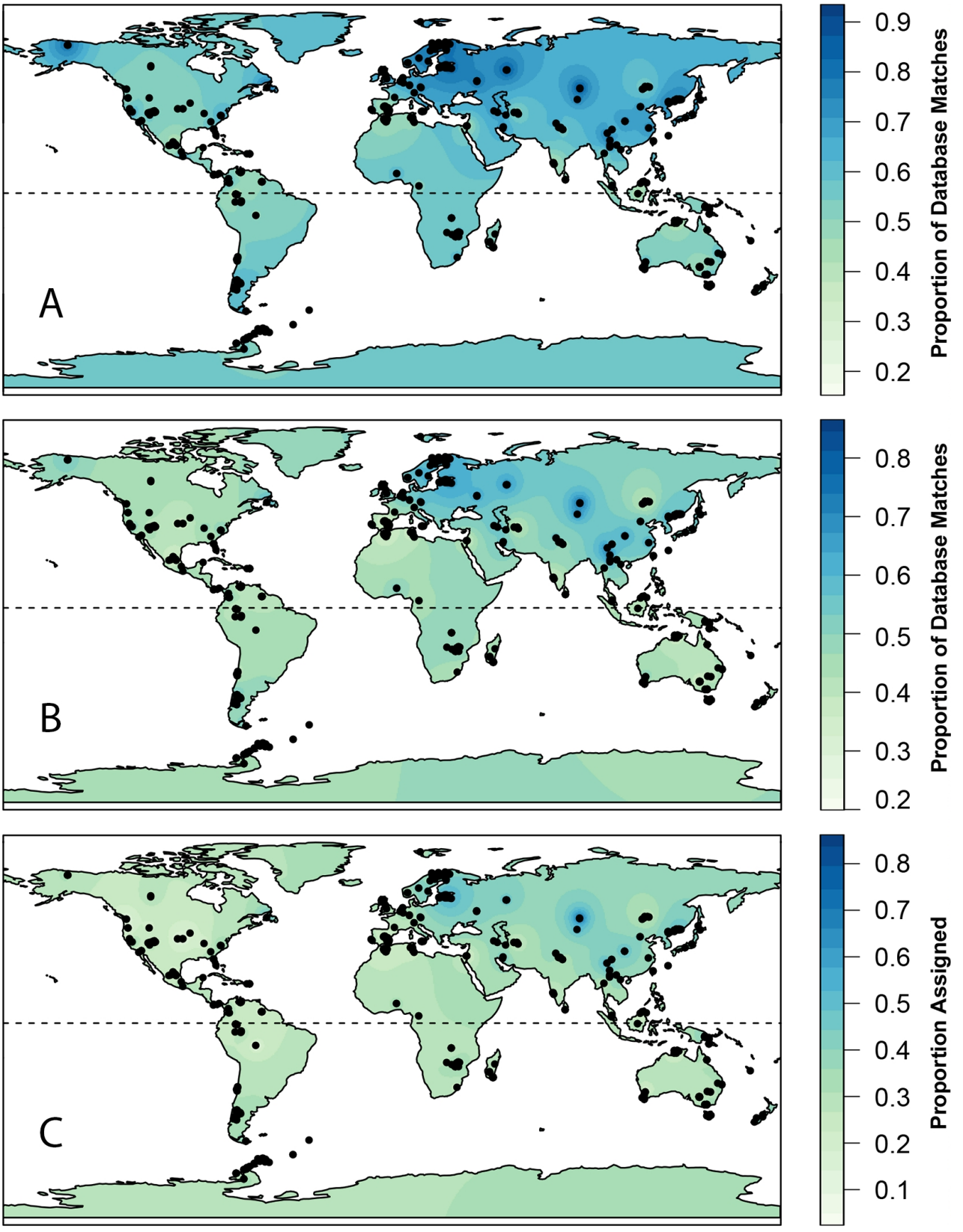
Figure 2. A barplot representing proportion of assigned fungal OTUs across 14 different terrestrial biomes globally in NCBI (green), UNITE (orange) and RDP (phylum level: blue; genus level: purple) databases.

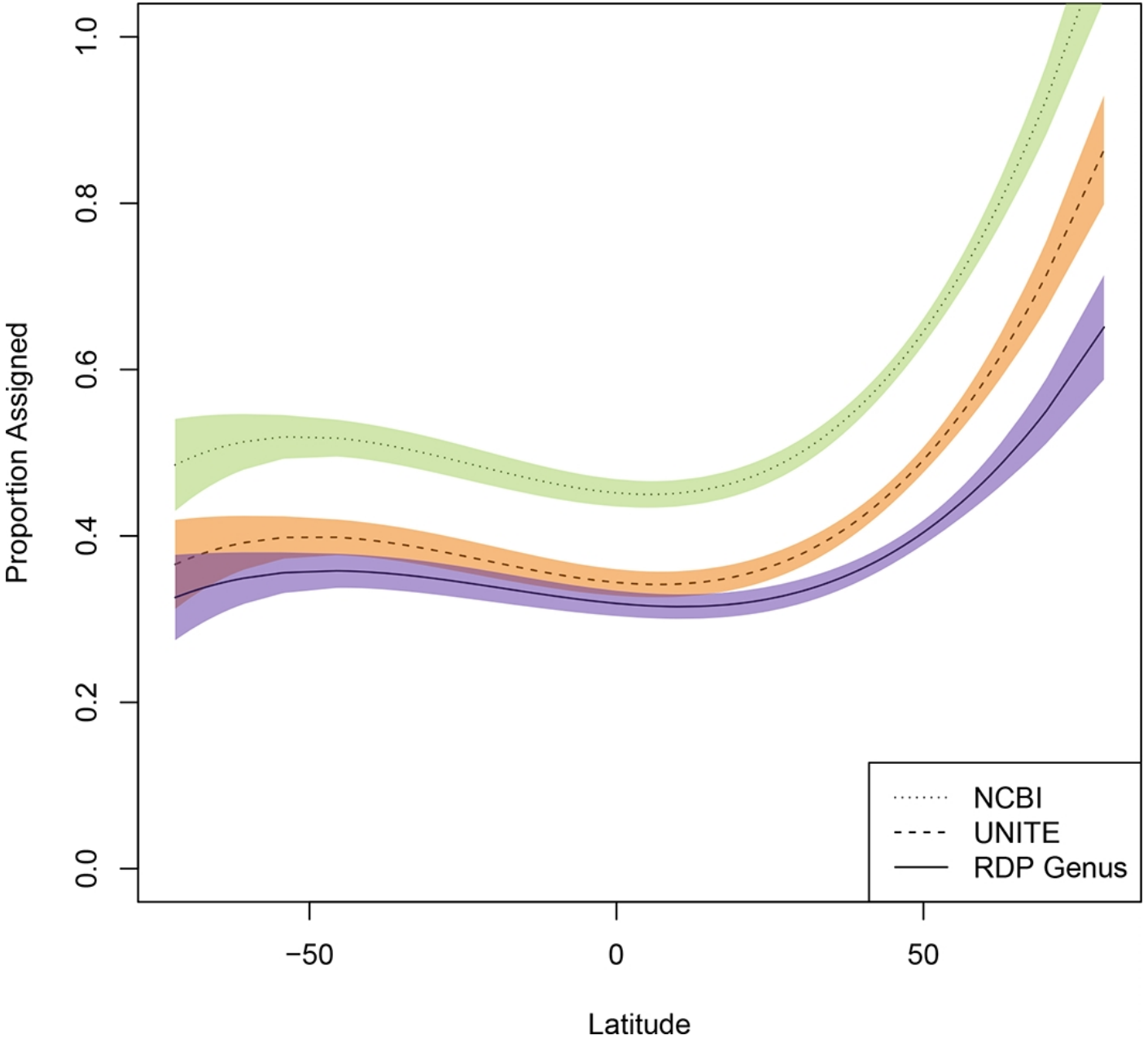
Figure 3. Interpolated database representation worldwide using the IDW algorithm for the **(A)** NCBI, **(B)** UNITE databases and **(C)** genus-level RDP assignments. Dark colours represent areas with a higher proportion of OTUs represented in the reference database. Points represent the 411 locations on which the interpolation is based.

Figure 4. Predicted fits and 95% confidence intervals for the best models (third-degree polynomial fits of latitude) explaining database representation for both the NCBI and UNITE databases, and genus-level RDP assignments.


NCBI
UNITE
RDPphylum80
RDPgenus80
Substrate
Wood
Soil
Skin
Sediment
Root
Moss
Marine
Litter
Leaf
Insect
Freshwater
Driftwood
Built_environment
Air
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Proportion OTUs Assigned


NCBI
UNITE
RDPphylum80
RDPgenus80
Biome
Tropical_montane_forests
Temperate_deciduous_forests
Temperate_coniferous_forests
Subantarctic
Southern_temperate_forests
Savannas
Moist_tropical_forests
Mediterranean
Maritime_antarctic
Mangrove
Grassland_and_shrubland
Dunes
Dry_tropical_forests
Boreal_forests
Arctic_tundra
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Proportion OTUs Assigned


A
B
C
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Proportion of Database Matches
Proportion of Database Matches
Proportion Assigned


Proportion Assigned
0.0
0.2
0.4
0.6
0.8
1.0
−50
0
50
Latitude
NCBI
UNITE
RDP Genus

Supplementary Information

Fig. S1. A treemap of 196 790 fungal OTUs across 14 different substrates. The area is proportional to the number of fungal OTUs for each substrate.

Fig. S2. A cross plot representing mean and standard deviation values of fungal OTUs for each substrate with a query coverage > 70% and sequence similarity > 70% in NCBI. Numbers in brackets indicate number of studies analysed for this substrate.

Fig. S3. A cross plot representing mean and standard deviation values of fungal OTUs for each substrate with a query coverage > 70% and sequence similarity > 70% in UNITE. Numbers in brackets indicate number of studies analysed for this substrate.

Fig. S4. A matplot representing a proportion of successfully assigned fungal OTUs across 14 different substrates identified with > 80% confidence at a given taxonomic level using the RDP classifier.

Fig. S5. A matplot representing a proportion of successfully assigned Ascomycota OTUs across 14 different substrates identified with > 80% confidence at a given taxonomic level using the RDP classifier.

Fig. S6. A matplot representing a proportion of successfully assigned Basidiomycota OTUs across 14 different substrates identified with > 80% confidence at a given taxonomic level using the RDP classifier.

Figure S7. 95% confidence interval plot for the IDW interpolation of database representation in the NCBI database.

Figure S8. 95% confidence interval plot for the IDW interpolation of database representation in the UNITE database.

Figure S9. 95% confidence interval plot for the IDW interpolation of successful taxonomic

assignment at the genus level using the RDP classifier and Warcup Fungal ITS training set 2.

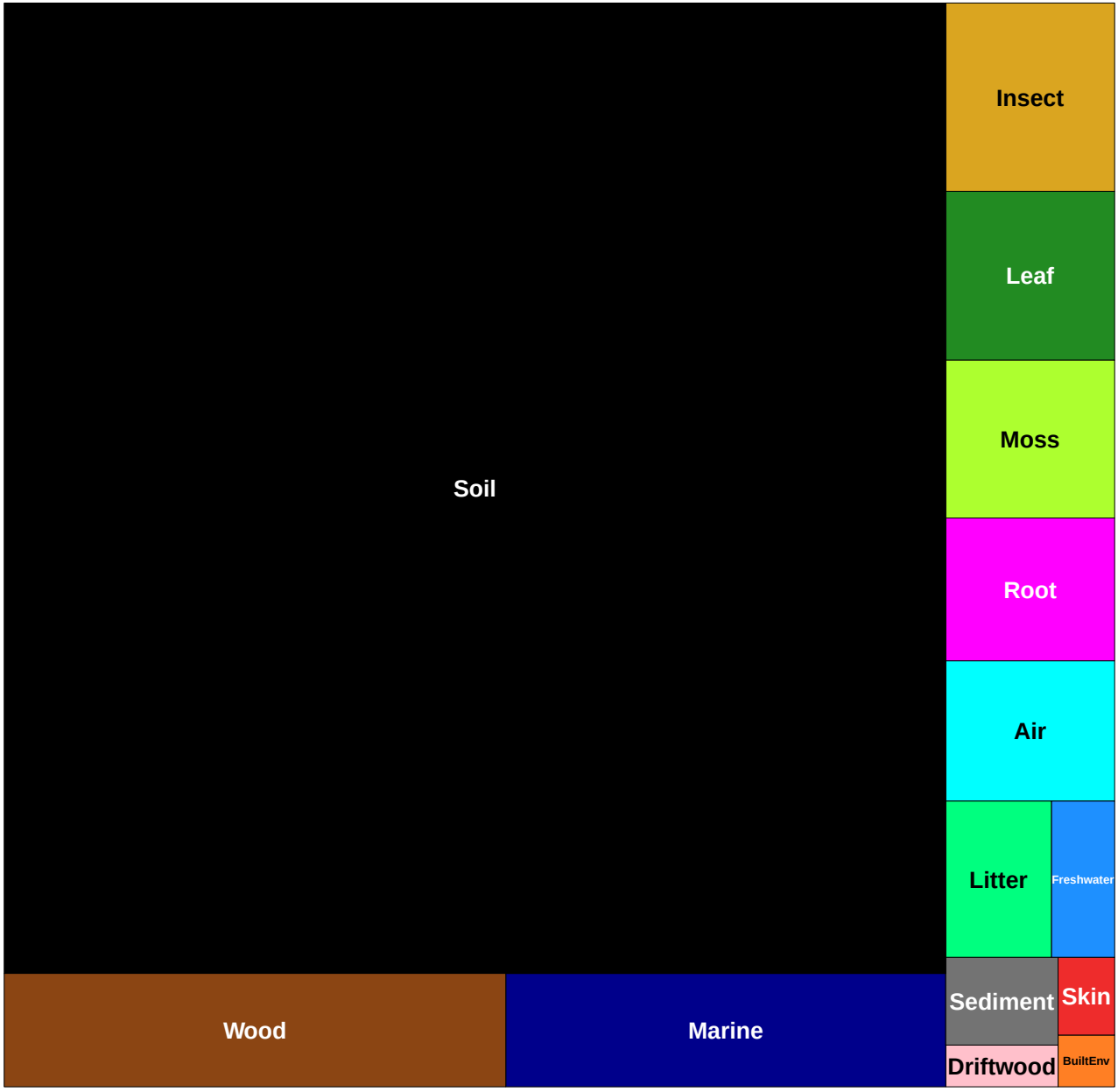
Table S1. Overview of studies included in a global meta-analysis of fungal communities across 14 different substrates.

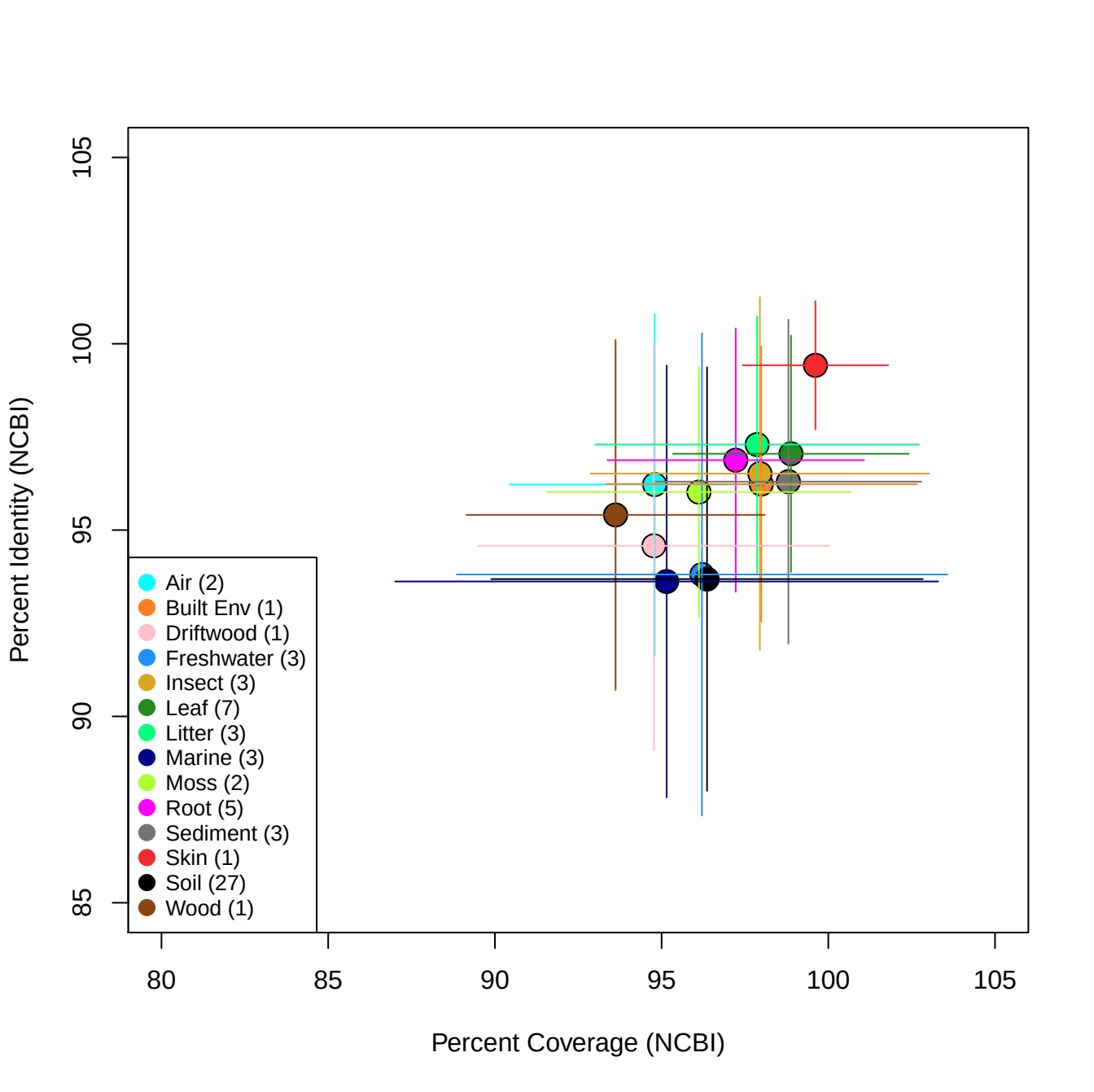
Table S2. Site, project, citation, and bioinformatic processing data for all localities in the global soil dataset.

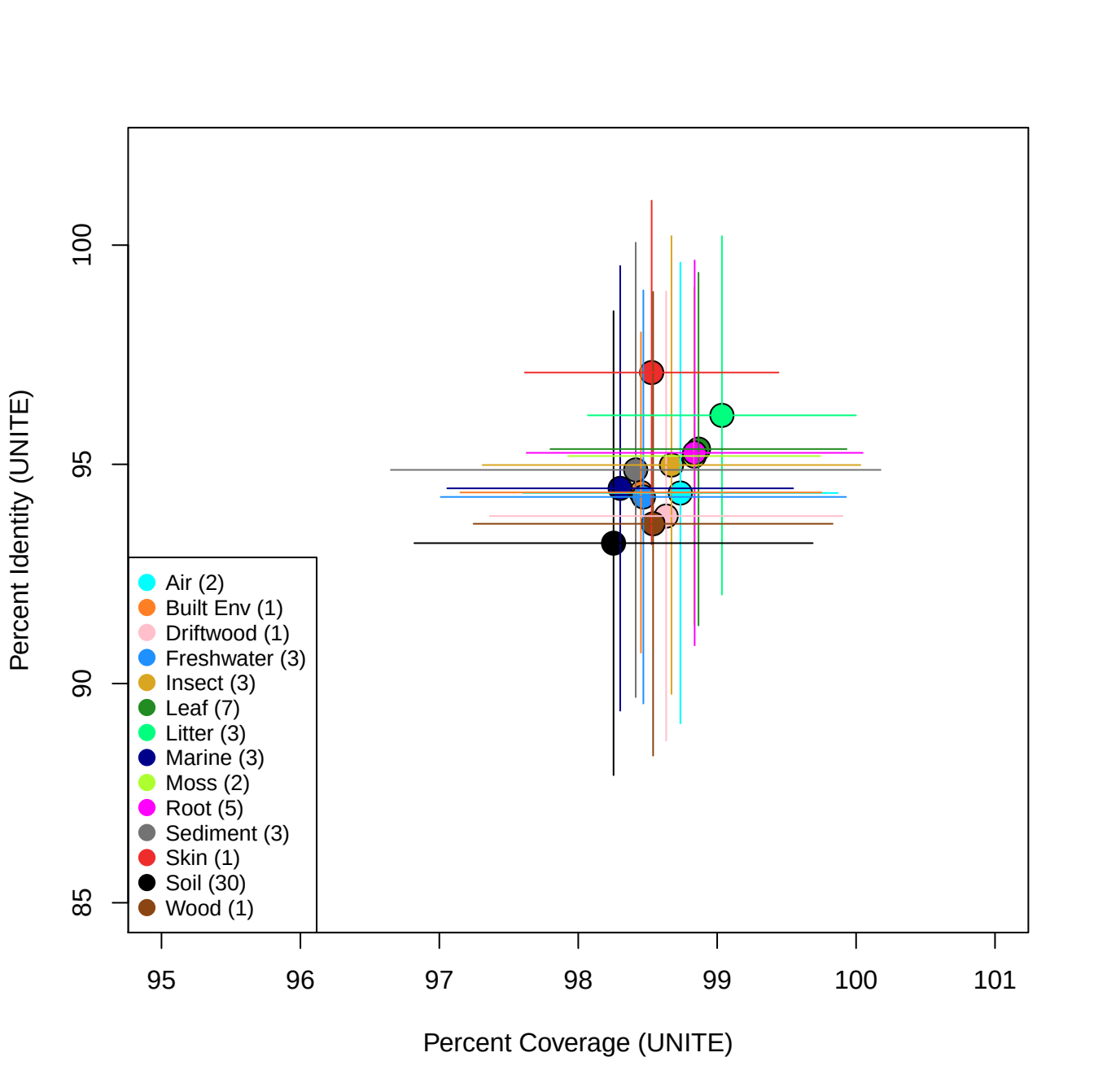
Appendix 1. A total dataset of fungal OTUs across 14 different substrates used in the study.

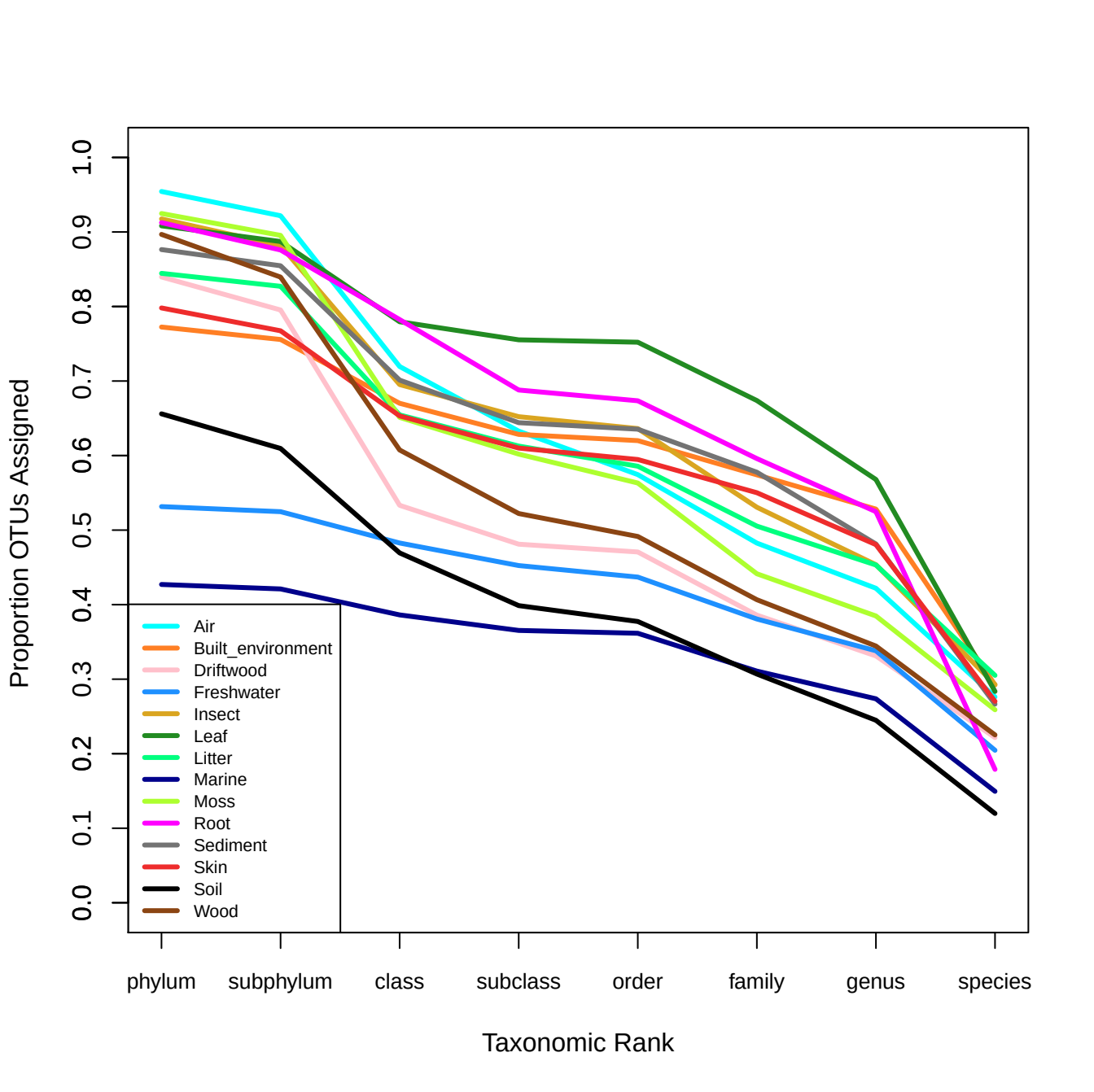
Appendix 2. A subset of fungal OTUs from soil representing 14 different biomes globally.

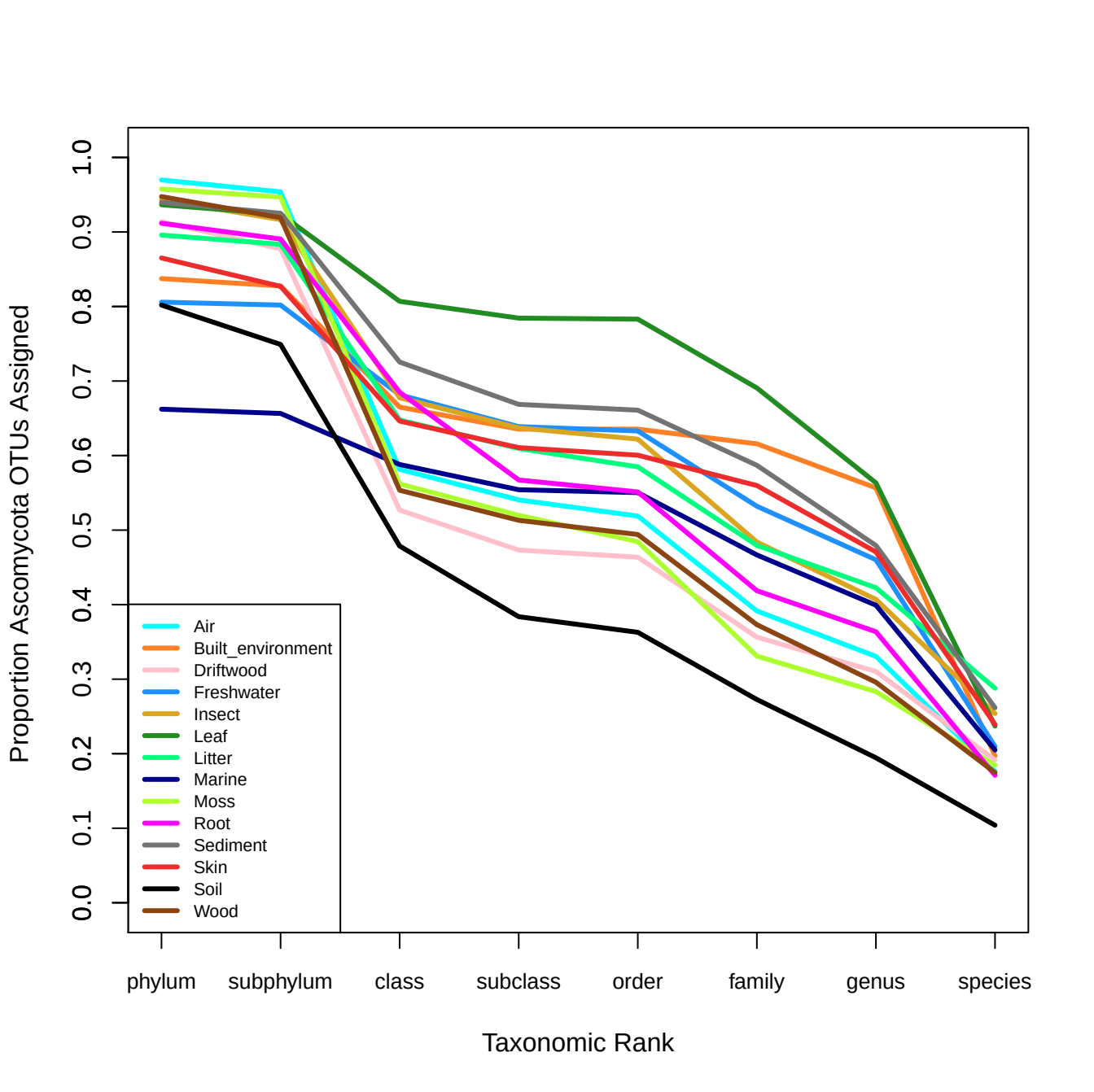
Distribution of fungal OTUs

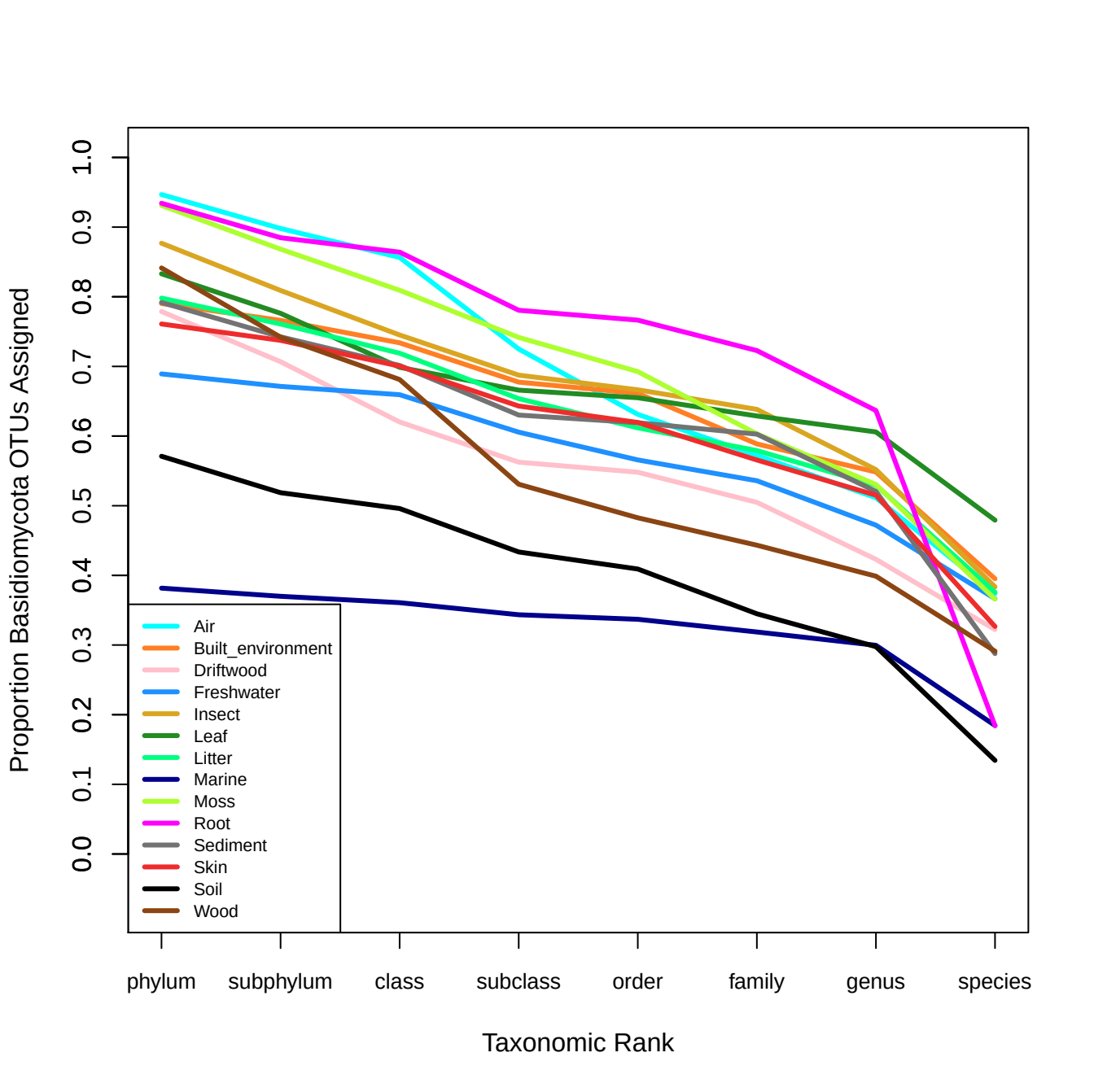


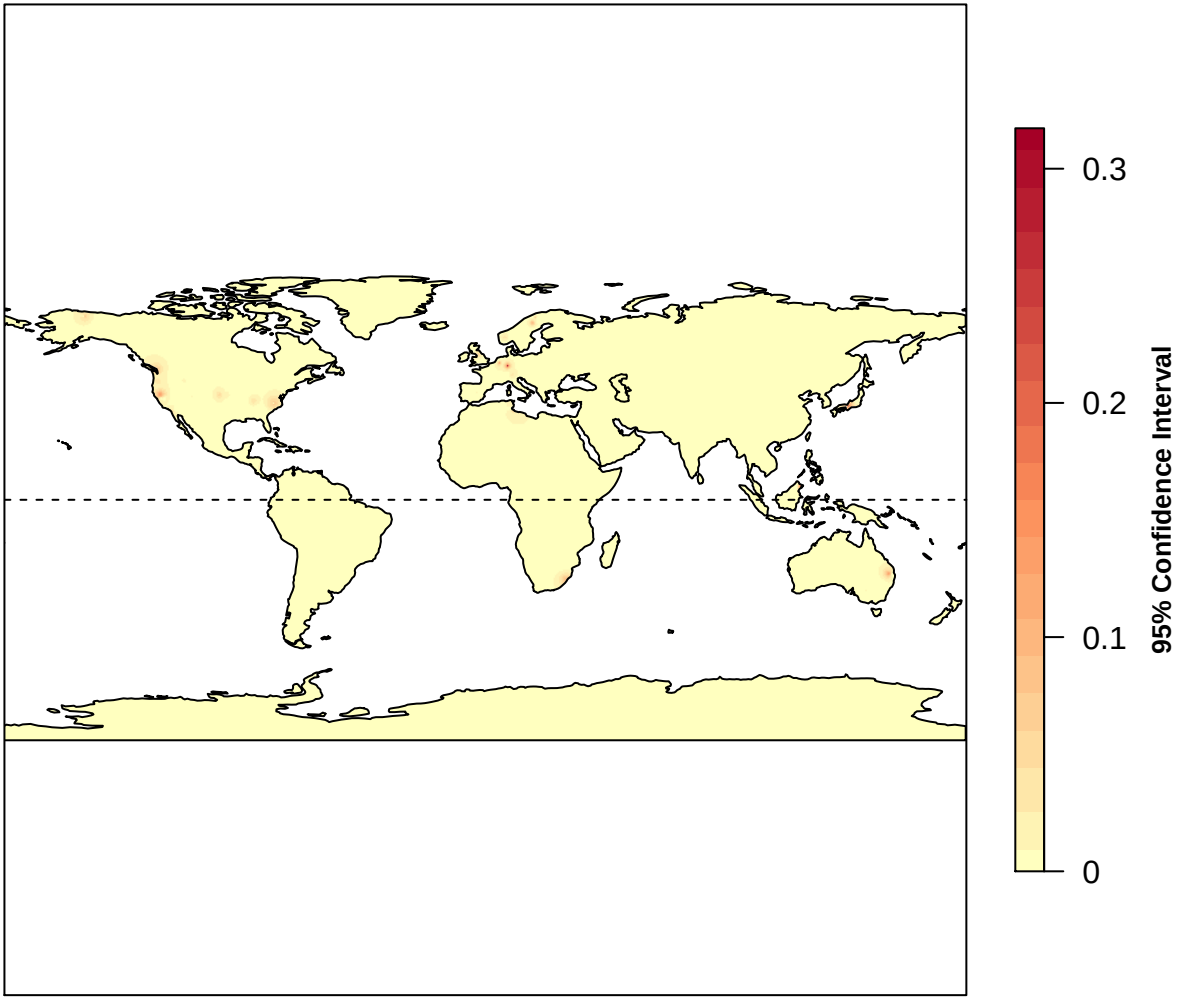

Percent Identity (NCBI)
105
100
95
90
85
80
85
90
95
100
105
Percent Coverage (NCBI)
Air (2)
Built Env (1)
Driftwood (1)
Freshwater (3)
Insect (3)
Leaf (7)
Litter (3)
Marine (3)
Moss (2)
Root (5)
Sediment (3)
Skin (1)
Soil (27)
Wood (1)

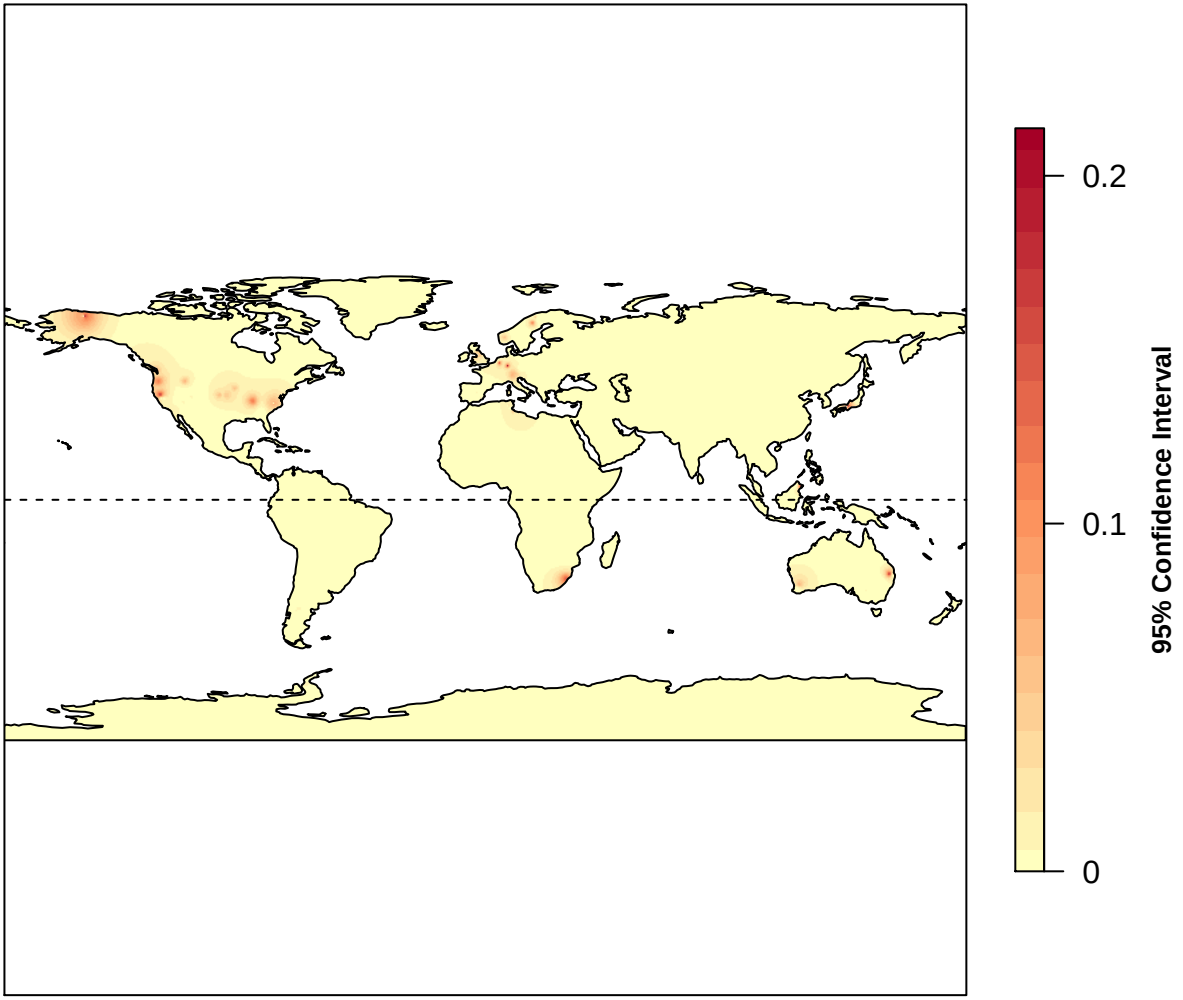

Percent Identity (UNITE)
100
95
90
85
95
96
97
98
99
100
101
Percent Coverage (UNITE)
Air (2)
Built Env (1)
Driftwood (1)
Freshwater (3)
Insect (3)
Leaf (7)
Litter (3)
Marine (3)
Moss (2)
Root (5)
Sediment (3)
Skin (1)
Soil (30)
Wood (1)


Proportion OTUs Assigned
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Air
Built_environment
Driftwood
Freshwater
Insect
Leaf
Litter
Marine
Moss
Root
Sediment
Skin
Soil
Wood
phylum
subphylum
class
subclass
order
family
genus
species
Taxonomic Rank


Proportion Ascomycota OTUs Assigned
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Air
Built_environment
Driftwood
Freshwater
Insect
Leaf
Litter
Marine
Moss
Root
Sediment
Skin
Soil
Wood
phylum
subphylum
class
subclass
order
family
genus
species
Taxonomic Rank


Proportion Basidiomycota OTUs Assigned
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Air
Built_environment
Driftwood
Freshwater
Insect
Leaf
Litter
Marine
Moss
Root
Sediment
Skin
Soil
Wood
phylum
subphylum
class
subclass
order
family
genus
species
Taxonomic Rank


0.3
0.2
0.1
0
95% Confidence Interval


0.2
0.1
0
95% Confidence Interval

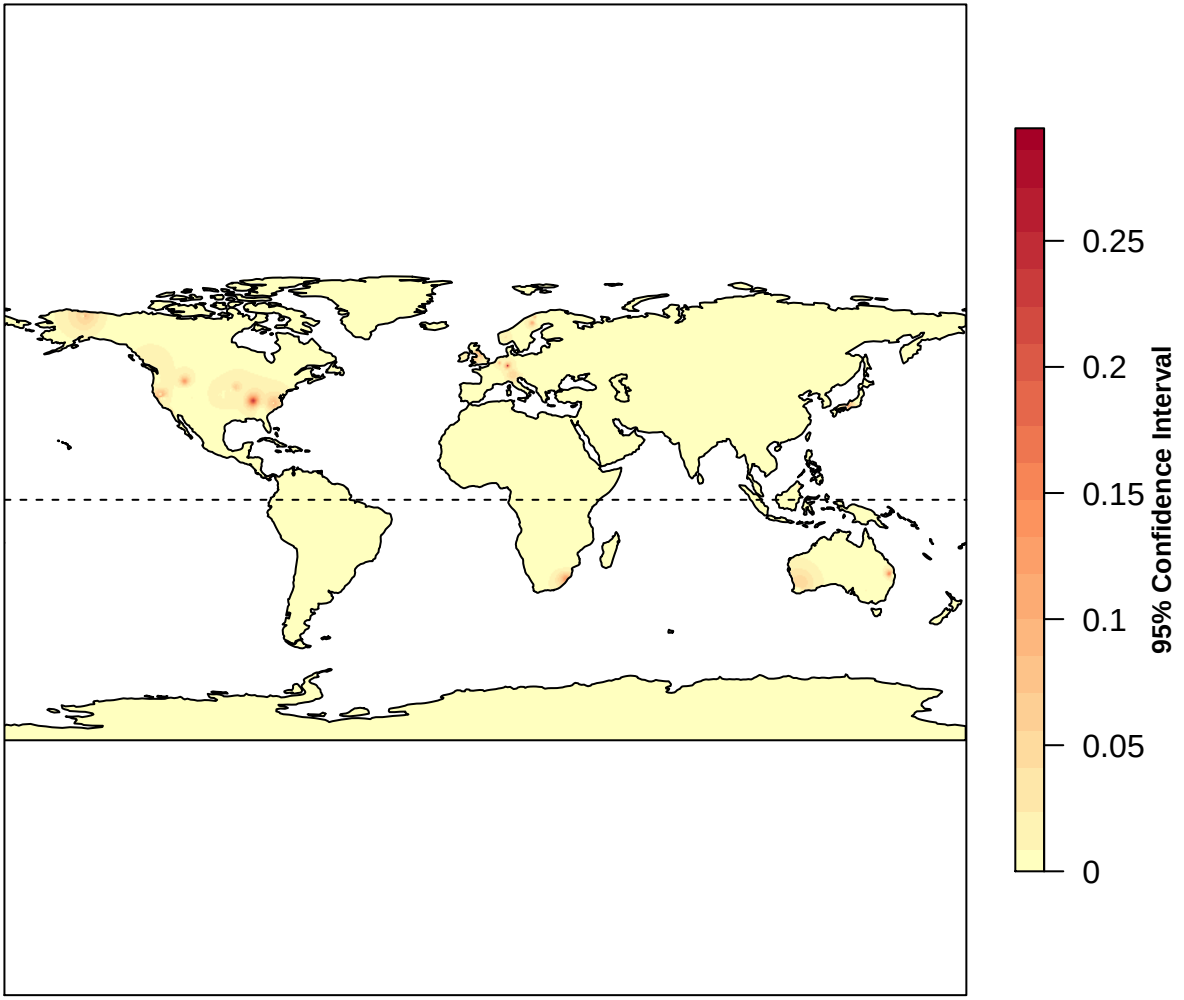

0.25
0.2
0.15
0.1
0.05
0
95% Confidence Interval

Table S1. Overview of studies included in a global meta-analysis of fungal communities across 14 different substrates.

No.	Project	Substrate	No. OTUs (> 99 nt)	pident NCBI	pident UNITE	RDP phylum80	RDP genus80	Reference
1	Aas_avenella_leaf	Leaf	198	0.58	0.45	0.97	0.51	unpublished
2	Aas_finse_root	Root	339	0.69	0.53	0.86	0.42	unpublished
3	Aas_finse_soil	Soil	315	0.65	0.51	0.75	0.36	unpublished
4	Adams_indoor_air	Built environment	479	0.50	0.27	0.77	0.53	Adams *et al.* (2013a)
5	Adams_outdoor_air	Air	404	0.45	0.24	0.86	0.58	Adams *et al.* (2013a)
6	Adams_skin	Skin	718	0.94	0.68	0.80	0.48	Adams *et al.* (2013b)
7	Arfi_new_caledonia_soil	Soil	36	0.65	0.59	0.94	0.64	Arfi *et al.* (2012)
8	Baerdsdatter_marine_fungi	Marine	1602	0.63	0.90	0.10	0.07	unpublished
9	Balint_meadow_soil	Soil	1266	0.50	0.40	0.88	0.28	Bálint *et al.* (2014)
10	Balint_poplar_leaf	Leaf	182	0.82	0.64	0.98	0.75	Bálint *et al.* (2013)
11	Barberan_panama_soil	Soil	20613	0.17	0.44	0.37	0.08	Barberán *et al.* (2015)
12	Bistorta_BioGeo	Root	997	0.63	0.45	0.91	0.44	Blaalid *et al.* (2014)
13	Blaalid_finse_soil	Soil	270	0.71	0.58	0.75	0.40	Blaalid *et al.* (2013)
14	Clemmensen_sweden_soil	Soil	1617	0.70	0.51	0.85	0.38	Clemmensen *et al.* (2013)
15	Cordier_beech_endophyte	Leaf	1028	0.96	0.45	0.85	0.41	Cordier *et al.* (2012)
16	Cox_antarctic_soil	Soil	599	0.53	0.43	0.70	0.29	Cox *et al.* (2016)
17	Davey_bryophyte_gradient	Moss	1751	0.54	0.42	0.95	0.35	Davey *et al.* (2013)
18	Davey_bryophyte_seasonal	Moss	2610	0.44	0.41	0.91	0.41	Davey *et al.* (2012)
19	De_Beeck_belgium_soil	Soil	177	0.77	0.68	0.90	0.49	De Beeck *et al.* (2014)
20	Duarte_leaf_litter	Litter	1044	0.74	0.63	0.85	0.48	Duarte *et al.* (2015)
21	Geml_netherlands_soil	Soil	3021	0.56	0.43	0.70	0.35	Geml *et al.* (2014)
22	Geml_arctic_alaska_soil	Soil	5358	0.70	0.28	0.86	0.30	Geml *et al.* (2016)

No.	Project	Substrate	No. OTUs (> 99 nt)	pident NCBI	pident UNITE	RDP phylum80	RDP genus80	Reference
23	Glacier_forefront_root	Root	636	0.62	0.48	0.85	0.56	Davey *et al.* (2015)
24	Ihrmark_sweden_soil	Soil	282	0.77	0.59	0.87	0.37	Ihrmark *et al.* (2012)
25	Jacobsen_insect	Insect	3511	0.62	0.47	0.92	0.46	Jacobsen *et al.* (2017)
26	Jeffries_marine_fungi	Marine	5939	0.33	0.28	0.51	0.32	Jeffries *et al.* (2016)
27	Jumpponen_usa_kansas_soil	Soil	519	excluded	0.44	0.65	0.21	Jumpponen *et al.* (2010)
28	Kadowaki_japan_soil	Soil	287	0.46	0.32	0.66	0.30	Kadowaki *et al.* (2014)
29	Kemler_eucalyptus_leaf	Leaf	1143	0.30	0.28	0.94	0.66	Kemler *et al.* (2013)
30	Kerfahi_borneo_soil	Soil	3934	0.30	0.23	0.76	0.29	Kerfahi *et al.* (2014)
31	Khomich_freshwater	Freshwater	300	0.73	0.71	0.57	0.41	Khomich *et al.* (2017)
32	Kostovcik_insect	Insect	367	0.74	0.62	0.96	0.55	Kostovcik *et al.* (2015)
33	Leff_global_grassland_soil	Soil	1430	0.39	0.36	0.28	0.17	Leff *et al.* (2015)
34	Li_marine_sediment	Sediment	809	0.53	0.41	0.89	0.48	Li *et al.* (2016)
35	Maestre_global_dryland_soil	Soil	23336	0.23	0.15	0.84	0.27	Maestre *et al.* (2015)
36	Mello_france_soil	Soil	272	0.73	0.62	0.79	0.35	Mello *et al.* (2011)
37	Miller_insect	Insect	1323	0.62	0.50	0.90	0.42	Miller *et al.* (2016)
38	Monard_sweden_soil	Soil	137	excluded	0.70	0.88	0.47	Monard *et al.* (2013)
39	Mueller_amazon_deforest_soil	Soil	15283	0.44	0.46	0.31	0.03	Mueller *et al.* (2014)
40	Mundra_root_spatial	Root	710	0.46	0.37	0.92	0.49	Mundra *et al.* (2015a)
41	Mundra_root_temporal	Root	1263	0.70	0.41	0.96	0.62	Mundra *et al.* (2015b)
42	Newsham_antarctic_soil	Soil	1936	0.32	0.27	0.70	0.27	Newsham *et al.* (2016)
43	Norden_air_spores	Air	3471	0.61	0.44	0.97	0.40	unpublished
44	Norden_wood_sawdust	Wood	9299	0.47	0.34	0.90	0.34	unpublished
45	Oliver_usa_georgia_soil	Soil	21488	0.49	0.32	0.75	0.33	Oliver *et al.* (2015)
46	Orgiazzi_italy_soil	Soil	164	excluded	0.74	0.98	0.66	Orgiazzi *et al.* (2012)

No.	Project	Substrate	No. OTUs (> 99 nt)	pident NCBI	pident UNITE	RDP phylum80	RDP genus80	Reference
47	Porter_canada_wetland_soil	Soil	2433	0.48	0.37	0.64	0.41	Porter *et al.* (2016)
48	Rama_driftwood	Driftwood	767	0.45	0.35	0.84	0.33	Rämä *et al.* (2014)
49	Shi_china_lat_gradient_soil	Soil	1000	0.68	0.59	0.80	0.54	Shi *et al.* (2014)
50	Siddique_beech_endophyte	Leaf	412	0.67	0.45	0.79	0.51	Siddique & Unterseher (2016)
51	Song_freshwater_Biwa	Freshwater	683	0.38	0.32	0.53	0.39	Song *et al.* (2017)
52	Tao_Zhang_arctic_endophytes	Leaf	250	0.54	0.36	0.95	0.39	Zhang & Yao (2015)
53	Tao_Zhang_arctic_freshwater	Freshwater	635	0.40	0.25	0.52	0.25	Zhang *et al.* (2016a)
54	Tao_Zhang_arctic_sediment	Sediment	107	0.73	0.49	0.79	0.39	Zhang *et al.* (2015)
55	Taylor_coast_marine_plankton	Marine	615	0.64	0.57	0.45	0.34	Taylor & Cunliffe (2016)
56	Tedersoo_global_soil	Soil	28908	0.36	0.26	0.81	0.36	Tedersoo *et al.* (2014)
57	Toju_japan_soil	Soil	1218	0.45	0.33	0.80	0.32	Toju *et al.* (2016)
58	Tripathi_brunei_soil	Soil	9206	0.13	0.30	0.39	0.08	Tripathi *et al.* (2016)
59	Troll_Soil	Soil	1006	0.57	0.33	0.52	0.26	unpublished
60	Unterseher_leaf	Leaf	1450	0.46	0.26	0.94	0.64	Unterseher *et al.* (2016)
61	Urbina_puerto_rico_soil	Soil	2354	0.28	0.24	0.79	0.25	Urbina *et al.* (2016)
62	Voriskova_oak_leaf_litter	Litter	628	0.55	0.38	0.84	0.44	Voříšková & Baldrian (2013)
63	X.Zhang_marine_sediment	Sediment	694	0.70	0.56	0.87	0.50	Zhang *et al.* (2016b)
64	Zifcakova_forest_litter	Litter	1029	0.69	0.52	0.84	0.43	Žifčáková *et al.* (2016)
65	Zifcakova_forest_soil	Soil	932	0.74	0.60	0.80	0.47	Žifčáková *et al.* (2016)
			196 790					

References:

Adams RI, Miletto M, Taylor JW & Bruns TD (2013a) Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *The ISME journal* **7**: 1262-1273.

Adams RI, Miletto M, Taylor JW & Bruns TD (2013b) The diversity and distribution of fungi on residential surfaces. *PloS one* **8**: e78866.

Arfi Y, Marchand C, Wartel M & Record E (2012) Fungal diversity in anoxic-sulfidic sediments in a mangrove soil. *Fungal Ecology* **5**: 282-285.

Bálint M, Schmidt PA, Sharma R, Thines M & Schmitt I (2014) An Illumina metabarcoding pipeline for fungi. *Ecology and evolution* **4**: 2642-2653.

Bálint M, Tiffin P, Hallström B, O'Hara RB, Olson MS, Fankhauser JD, Piepenbring M & Schmitt I (2013) Host genotype shapes the foliar fungal microbiome of balsam poplar (Populus balsamifera). *PLoS One* **8**: e53987.

Barberán A, McGuire KL, Wolf JA, Jones FA, Wright SJ, Turner BL, Essene A, Hubbell SP, Faircloth BC & Fierer N (2015) Relating belowground microbial composition to the taxonomic, phylogenetic, and functional trait distributions of trees in a tropical forest. *Ecology letters* **18**: 1397-1405.

Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk P & Kauserud H (2013) ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources* **13**: 218-224.

Blaalid R, Davey ML, Kauserud H, Carlsen T, Halvorsen R, Høiland K & Eidesen PB (2014) Arctic root-associated fungal community composition reflects environmental filtering. *Molecular ecology* **23**: 649-659.

Clemmensen K, Bahr A, Ovaskainen O, Dahlberg A, Ekblad A, Wallander H, Stenlid J, Finlay R, Wardle D & Lindahl B (2013) Roots and associated fungi drive long-term carbon sequestration in boreal forest. *Science* **339**: 1615-1618.

Cordier T, Robin C, Capdevielle X, Desprez-Loustau M-L & Vacher C (2012) Spatial variability of phyllosphere fungal assemblages: genetic distance predominates over geographic distance in a European beech stand (Fagus sylvatica). *Fungal ecology* **5**: 509-520.

Cox F, Newsham KK, Bol R, Dungait JA & Robinson CH (2016) Not poles apart: Antarctic soil fungal communities show similarities to those of the distant Arctic. *Ecology letters*.

Davey M, Blaalid R, Vik U, Carlsen T, Kauserud H & Eidesen PB (2015) Primary succession of Bistorta vivipara (L.) Delabre (Polygonaceae) root-associated fungi mirrors plant succession in two glacial chronosequences. *Environmental Microbiology* **17**: 2777-2790.

Davey ML, Heegaard E, Halvorsen R, Ohlson M & Kauserud H (2012) Seasonal trends in the biomass and structure of bryophyte-associated fungal communities explored by 454 pyrosequencing. *New Phytologist* **195**: 844-856.

Davey ML, Heegaard E, Halvorsen R, Kauserud H & Ohlson M (2013) Amplicon-pyrosequencing-based detection of compositional shifts in bryophyte-associated fungal communities along an elevation gradient.

Molecular ecology **22**: 368-383.

De Beeck MO, Lievens B, Busschaert P, Declerck S, Vangronsveld J & Colpaert JV (2014) Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies. *PLoS One* **9**: e97629.

Duarte S, Bärlocher F, Trabulo J, Cássio F & Pascoal C (2015) Stream-dwelling fungal decomposer communities along a gradient of eutrophication unraveled by 454 pyrosequencing. *Fungal Diversity* **70**: 127-148.

Geml J, Semenova TA, Morgado LN & Welker JM (2016) Changes in composition and abundance of functional groups of arctic fungi in response to long-term summer warming. *Biology Letters* **12**: 20160503.

Geml J, Gravendeel B, van der Gaag KJ, Neilen M, Lammers Y, Raes N, Semenova TA, de Knijff P & Noordeloos ME (2014) The contribution of DNA metabarcoding to fungal conservation: diversity assessment, habitat partitioning and mapping red-listed fungi in protected coastal Salix repens communities in the Netherlands. *PloS one* **9**: e99852.

Ihrmark K, Bödeker IT, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J, Brandström-Durling M & Clemmensen KE (2012) New primers to amplify the fungal ITS2 region–evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology* **82**: 666-677.

Jacobsen RM, Kauserud H, Sverdrup-Thygeson A, Bjorbækmo MM & Birkemoe T (2017) Wood-inhabiting insects can function as targeted vectors for decomposer fungi. *Fungal Ecology* **29**: 76-84.

Jeffries TC, Curlevski NJ, Brown MV, Harrison DP, Doblin MA, Petrou K, Ralph PJ & Seymour JR (2016) Partitioning of fungal assemblages across different marine habitats. *Environmental microbiology reports*.

Jumpponen A, Jones KL & Blair J (2010) Vertical distribution of fungal communities in tallgrass prairie soil. *Mycologia*.

Kadowaki K, Sato H, Yamamoto S, Tanabe AS, Hidaka A & Toju H (2014) Detection of the horizontal spatial structure of soil fungal communities in a natural forest. *Population ecology* **56**: 301-310.

Kemler M, Garnas J, Wingfield MJ, Gryzenhout M, Pillay K-A & Slippers B (2013) Ion Torrent PGM as tool for fungal community analysis: a case study of endophytes in Eucalyptus grandis reveals high taxonomic diversity. *PLoS One* **8**: e81718.

Kerfahi D, Tripathi BM, Lee J, Edwards DP & Adams JM (2014) The impact of selective-logging and forest clearance for oil palm on fungal communities in Borneo. *PloS one* **9**: e111525.

Khomich M, Davey ML, Kauserud H, Rasconi S & Andersen T (2017) Fungal communities in Scandinavian lakes along a longitudinal gradient. *Fungal Ecology* **27**: 36-46.

Kostovcik M, Bateman CC, Kolarik M, Stelinski LL, Jordal BH & Hulcr J (2015) The ambrosia symbiosis is specific in some species and promiscuous in others: evidence from community pyrosequencing. *The ISME journal* **9**: 126-138.

Leff JW, Jones SE, Prober SM, Barberán A, Borer ET, Firn JL, Harpole WS, Hobbie SE, Hofmockel KS

& Knops JM (2015) Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proceedings of the National Academy of Sciences* **112**: 10967-10972.

Li W, Wang MM, Wang XG, Cheng XL, Guo JJ, Bian XM & Cai L (2016) Fungal communities in sediments of subtropical Chinese seas as estimated by DNA metabarcoding. *Scientific reports* **6**.

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

Mello A, Napoli C, Murat C, Morin E, Marceddu G & Bonfante P (2011) ITS-1 versus ITS-2 pyrosequencing: a comparison of fungal populations in truffle grounds. *Mycologia* **103**: 1184-1193.

Miller KE, Hopkins K, Inward DJ & Vogler AP (2016) Metabarcoding of fungal communities associated with bark beetles. *Ecology and evolution*.

Monard C, Gantner S & Stenlid J (2013) Utilizing ITS1 and ITS2 to study environmental fungal diversity using pyrosequencing. *FEMS microbiology ecology* **84**: 165-175.

Mueller RC, Paula FS, Mirza BS, Rodrigues JL, Nüsslein K & Bohannan BJ (2014) Links between plant and fungal communities across a deforestation chronosequence in the Amazon rainforest. *The ISME journal* **8**: 1548-1550.

Mundra S, Halvorsen R, Kauserud H, Müller E, Vik U & Eidesen PB (2015a) Arctic fungal communities associated with roots of Bistorta vivipara do not respond to the same fine-scale edaphic gradients as the aboveground vegetation. *New Phytologist* **205**: 1587-1597.

Mundra S, Bahram M, Tedersoo L, Kauserud H, Halvorsen R & Eidesen PB (2015b) Temporal variation of Bistorta vivipara-associated ectomycorrhizal fungal communities in the High Arctic. *Molecular ecology* **24**: 6289-6302.

Newsham KK, Hopkins DW, Carvalhais LC, Fretwell PT, Rushton SP, O'Donnell AG & Dennis PG (2016) Relationship between soil fungal diversity and temperature in the maritime Antarctic. *Nature Climate Change* **6**: 182.

Oliver AK, Callaham MA & Jumpponen A (2015) Soil fungal communities respond compositionally to recurring frequent prescribed burning in a managed southeastern US forest ecosystem. *Forest Ecology and Management* **345**: 1-9.

Orgiazzi A, Lumini E, Nilsson RH, Girlanda M, Vizzini A, Bonfante P & Bianciotto V (2012) Unravelling soil fungal communities from different Mediterranean land-use backgrounds. *PloS one* **7**: e34847.

Porter TM, Shokralla S, Baird D, Golding GB & Hajibabaei M (2016) Ribosomal DNA and plastid markers used to sample fungal and plant communities from wetland soils reveals complementary biotas. *PloS one* **11**: e0142759.

Rämä T, Nordén J, Davey ML, Mathiassen GH, Spatafora JW & Kauserud H (2014) Fungi ahoy! Diversity on marine wooden substrata in the high North. *Fungal Ecology* **8**: 46-58.

Shi L-L, Mortimer PE, Slik JF, Zou X-M, Xu J, Feng W-T & Qiao L (2014) Variation in forest soil

fungal diversity along a latitudinal gradient. *Fungal diversity* **64**: 305-315.

Siddique A & Unterseher M (2016) A cost-effective and efficient strategy for Illumina sequencing of fungal communities: A case study of beech endophytes identified elevation as main explanatory factor for diversity and community composition. *Fungal Ecology* **20**: 175-185.

Song P, Tanabe S, Yi R, Kagami M, Liu X & Ban S (2017) Fungal community structure at pelagic and littoral sites in Lake Biwa determined with high-throughput sequencing. *Limnology* 1-11.

Taylor JD & Cunliffe M (2016) Multi-year assessment of coastal planktonic fungi reveals environmental drivers of diversity and abundance. *The ISME journal*.

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

Toju H, Kishida O, Katayama N & Takagi K (2016) Networks Depicting the Fine-Scale Co-Occurrences of Fungi in Soil Horizons. *PloS one* **11**: e0165987.

Tripathi BM, Song W, Slik J, Sukri RS, Jaafar S, Dong K & Adams JM (2016) Distinctive Tropical Forest Variants Have Unique Soil Microbial Communities, But Not Always Low Microbial Diversity. *Frontiers in microbiology* **7**.

Unterseher M, Siddique AB, Brachmann A & Peršoh D (2016) Diversity and Composition of the Leaf Mycobiome of Beech (Fagus sylvatica) Are Affected by Local Habitat Conditions and Leaf Biochemistry. *PloS one* **11**: e0152878.

Urbina H, Scofield DG, Cafaro M & Rosling A (2016) DNA-metabarcoding uncovers the diversity of soil-inhabiting fungi in the tropical island of Puerto Rico. *Mycoscience* **57**: 217-227.

Voříšková J & Baldrian P (2013) Fungal community on decomposing leaf litter undergoes rapid successional changes. *The ISME journal* **7**: 477-486.

Zhang T & Yao Y-F (2015) Endophytic fungal communities associated with vascular plants in the high arctic zone are highly diverse and host-plant specific. *PloS one* **10**: e0130051.

Zhang T, Wang NF, Zhang YQ, Liu HY & Yu LY (2015) Diversity and distribution of fungal communities in the marine sediments of Kongsfjorden, Svalbard (High Arctic). *Scientific Reports* **5**: 14524.

Zhang T, Wang N-F, Zhang Y-Q, Liu H-Y & Yu L-Y (2016a) Diversity and distribution of aquatic fungal communities in the Ny-Ålesund region, Svalbard (High Arctic). *Microbial ecology* **71**: 543-554.

Zhang X-Y, Wang G-H, Xu X-Y, Nong X-H, Wang J, Amin M & Qi S-H (2016b) Exploring fungal diversity in deep-sea sediments from Okinawa Trough using high-throughput Illumina sequencing. *Deep Sea Research Part I: Oceanographic Research Papers* **116**: 99-105.

Žifčáková L, Větrovský T, Howe A & Baldrian P (2016) Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. *Environmental microbiology* **18**: 288-301.