Environmental DNA

WILEY

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

ORIGINAL ARTICLE OPEN ACCESS

Consistent Species Richness Patterns but Not Richness Estimates Based on Both ASV and OTU Inference Methods on ITS2-Based Soil Fungal Communities

Julia Kyaschenko[1] | Jeanette Tångrot[2] | Anders Dahlberg[3] | Björn D. Lindahl[4] | Anna Rosling[1]

[1]Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden | [2]Department of Plant Physiology, National Bioinformatics Infrastructure Sweden (NBIS), Umeå University, Umeå, Sweden | [3]Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden | [4]Department of Soil and Environment, Swedish University of Agricultural Sciences, Uppsala, Sweden

Correspondence: Julia Kyaschenko (julia.kyaschenko@ebc.uu.se)

Received: 11 August 2025 | **Revised:** 27 December 2025 | **Accepted:** 12 January 2026

Keywords: community dynamics | DNA barcoding | ecosystem | fungi | high-throughput nucleotide sequencing | molecular taxonomy | soil biodiversity | species specificity | taxonomy

ABSTRACT

Fungi play essential roles in key ecosystem functions and processes, yet they often occur in inconspicuous, species-rich, and complex communities that remain difficult to study. Studies of fungal communities based on DNA extracted from environmental samples commonly rely on clustering sequence reads into units of diversity, followed by taxonomic identification and, in some cases, linkage to ecological traits. In this study, we evaluated how two clustering approaches—amplicon sequence variants (ASVs) and operational taxonomic units (OTUs)—affect the characterization of fungal communities. Despite minor differences, both approaches recovered consistent taxonomic patterns and community structure. Although both methods produced a similar total number of sequence clusters, they differed in representation of fungal community composition. All ASV representative sequences matched OTU representative sequences with at least 92.2% similarity, whereas several rare OTUs showed low similarity to ASV reads, suggesting differences in the detection of low-abundance taxa. However, only a small fraction of OTU reads ($<0.1\%$) lacked a corresponding ASV, indicating that ASVs captured nearly all OTU-defined taxa. In contrast, 14% of ASV reads assigned to species hypotheses (SHs) did not match any OTU reads assigned to SHs, whereas only 1.3% of OTU SH-assigned reads lacked a corresponding ASV match. ASVs generally provided higher resolution than OTUs, as abundant SHs were often represented by multiple ASVs, suggesting that ASVs capture intraspecific diversity. Consequently, ASVs should not be used as direct species proxies but instead require post hoc grouping to reflect species-level diversity. OTUs-based community composition aligned more clearly with soil properties, particularly the N:C ratio. Overall, both approaches provided a similar overview of broad-scale species richness. The choice between two clustering methods depends on the research question and the desired level of taxonomic resolution, and our results provide little support for the claim that ASVs should categorically replace taxonomic units in marker-gene data analysis.

1 | Introduction

Fungi often exist in speciose, complex and diverse communities hidden in soils, wood, dead organic matter or live hosts, making them difficult to observe and study. The standard approach to overcome these challenges is the use of DNA metabarcoding of the nuclear ribosomal internal transcribed spacer (ITS) as an amplified marker, because it provides the species-level resolution often needed for fungal community studies (Schoch et al. 2012; Kauserud 2023). ITS variation within species is usually lower

---

https://doi.org/10.1002/edn3.70246

compared to variation between species, and clustering of fungal ITS sequences captures this barcoding gap to generate operational taxonomic units (OTUs; Blaxter et al. 2005). OTUs are commonly used as proxies for taxa in community studies, often adjusted to correspond to the species level (Kõljalg et al. 2020). DNA sequences are clustered into OTUs based on a similarity threshold, which typically varies between 97% and 99% depending on the clustering algorithm to strike a balance between taxonomic resolution on the one hand and intraspecific genetic variation and sequencing errors, on the other (Nilsson, Anslan, et al. 2019).

Intrinsic sequencing errors coupled with inadequate choice of clustering algorithms can lead to artificially inflated estimates of community richness. Single-linkage clustering of sequences with a predefined 99% similarity threshold implies that a sequence is at least 99% similar to at least one other sequence in the OTU (also referred to as nearest-neighbor clustering). In single-linkage clustering, OTUs will always be the same, irrespective of which sequences were used to initiate the clustering process, ensuring consistency and reproducibility in identifying OTUs (Lindahl et al. 2013). Moreover, in single-linkage methods, clusters expand until no similar sequences remain to be included, therefore enhancing their efficacy in handling sequencing errors while making efficient use of barcoding gaps to distinguish species (Huse et al. 2010). However, successful single-linkage clustering requires that OTUs are distinct from sister taxa with a clear barcoding gap; otherwise, there is a risk of merging sequences into overly extensive clusters including many taxa (Lindahl et al. 2013).

Starting in studies of prokaryotes, it has been suggested that communities should instead be analyzed based on exact nucleotide sequences, such as amplicon sequence variants (ASVs) as inferred, for example, from the software *DADA2* (Callahan et al. 2016). This software utilizes read quality information to control for sequencing error while preserving intraspecies variation. ASVs are claimed to retain only biologically relevant sequences, resolve fine-scale variation better, and retain fewer incorrect sequences than the best OTU method (Callahan et al. 2017; Estensmo et al. 2021). However, it has been argued that ASVs are conceptually similar to OTUs, since in practice *DADA2* is also based on clustering of sequences, although without a set similarity threshold (Kauserud 2023). Furthermore, ASVs as inferred by the software *DADA2* were originally developed for analyzing prokaryotic communities using 16S rRNA gene sequencing (Eren et al. 2013). Metabarcoding studies of prokaryote communities do not rely on barcoding gaps (i.e., larger interspecific than intraspecific variation), as they are not designed to delineate species-level taxonomic units (Frøslev et al. 2017), and therefore cannot reliably serve as proxies for fungal species (Furneaux et al. 2021; Eshghi Sahraei et al. 2022). Despite this concern, the use of ASVs has gained some popularity also in fungal community studies during the last decade (Krah and March-Salas 2022; Renaudin et al. 2023; Zhenting et al. 2024).

Fungal metabarcoding studies are usually based on short-read amplicons of ITS1 or ITS2 of the nuclear ribosomal operon (Kõljalg et al. 2013; Tedersoo et al. 2014). Fungal ITS sequences often contain a high degree of genetic variability also between closely related species, whereas intraspecific variation is considerably lower (Schoch et al. 2012). However, ITS amplicons also include PCR errors and sequencing artifacts, which may differently impact the bioinformatic output of ASV- and OTU-based methods. It is essential to understand how well ASVs and OTUs correspond to fungal species concepts and taxonomies used in curated reference databases, such as UNITE (Abarenkov et al. 2024; Kõljalg et al. 2005), to evaluate the reliability of downstream analyses and interpretations. In fungal systematics, the species concept is often operationalized through a combination of genetic similarity (typically 97%–99% ITS sequence similarity within species), ecological niche, and, when possible, morphology (Lücking et al. 2020). Since many fungal species remain undescribed or lack clear diagnostic features, molecular species hypotheses (like UNITE's Species Hypotheses, SHs) serve as species proxies (Nilsson, Larsson, et al. 2019). Investigation of how ASVs and OTUs capture fungal diversity in environmental DNA samples can therefore provide critical insights into the robustness and ecological relevance of metabarcoding-based community analyses, particularly to what extent the resulting units reflect real biological differences, such as species boundaries or ecological roles. In Illumina generated ITS2 amplicons of fungi, alpha and beta species richness and diversity metrics of leaf litter communities based on the two clustering approaches were strongly positively correlated across samples (Glassman and Martiny 2018). However, ASV richness appeared to be slightly less responsive to “site” and “inoculum” treatments than did OTU richness (Glassman and Martiny 2018), which could be related to the fact that the ASVs reflect more subtle genetic differences within species at the expense of broader-scale community shifts, and that fungal response traits are homogenized within species by genetic recombination. With long amplicon sequencing using PacBio, inference of ASVs with *DADA2* captured intra-specific genetic variation only for abundant taxa, while rare taxa were often lost compared to when various OTU clustering methods were applied. Together this resulted in higher ASV-based richness estimates compared to those inferred from other cluster-based methods (Eshghi Sahraei et al. 2022). This likely reflects the conservative nature of *DADA2*, which models and removes sequencing errors to capture minute differences but may filter out low-abundance variants as noise (Callahan et al. 2016). The likelihood that sequencing errors introduce new variants increases with amplicon length and may render ASVs unsuitable for longer amplicons. In contrast, OTU clustering methods group sequences based on similarity thresholds regardless of abundance, making them more inclusive of rare taxa, though at the risk of overestimating diversity by less efficient correction for sequencing artifacts (Edgar 2013; Nguyen et al. 2016; Frøslev et al. 2017). It has been proposed that the ASV approach may perform poorly for fungal groups with dikaryotic (Basidiomycota) or diploid (most unicellular fungal groups) genomes that commonly exhibit two or multiple different rRNA gene and ITS sequences per genome (Egan et al. 2018; Lindner et al. 2013; Runnel et al. 2024), and even more so in lineages where the ribosomal operon is not conserved in a tandem repeat, but exists in multiple independently evolving copies, as in the case of Glomeromycota (Maeda et al. 2018). If used as a proxy for species, ASV data are likely to overestimate richness for many fungal groups (due to haplotype variation), but underestimate richness by dismissing sequences from lowly abundant species as artifacts (see also Joos et al. 2020; Chiarello et al. 2022; Tedersoo et al. 2022; Eshghi Sahraei et al. 2022).

In this study, we compared the performance of bioinformatic pipelines clustering ITS2 reads either into OTUs by a single linkage algorithm at 99% similarity or into ASVs, and evaluated how this affected richness estimates of fungal community data from Lindahl et al. (2021) obtained from 359 plots from the Swedish Forest Soil Inventory. We compared the effectiveness of the two methods by assessing (1) the overlap between sequencing data retained and represented by each clustering method, (2) how well each method allowed accurate and comprehensive identification of fungal species against the UNITE database, and (3) how the resulting fungal communities obtained by either ASV- or OTU-based methods differed in terms of estimated richness and ecological relationships with an important environmental variable (soil N:C ratio).

2 | Materials and Methods

The study employed environmental and sequence data that was already acquired, analyzed, and published (see Lindahl et al. 2021) and is available at NCBI SRA as Bioproject PRJNA693127. A detailed description of sites, sampling, sequencing, and library preparation is provided in Lindahl et al. (2021). In short, sampling was carried out in connection with the Swedish Forest Soil Inventory (Fridman et al. 2014) during 2014–2016. Organic topsoil samples were collected in 3 m^2 plots in 359 conifer-dominated forest stands distributed across the Swedish boreal biome from 60° to 68° N and from 12° to 24° E. ITS2 markers were amplified from extracted DNA using the fungal-specific primer gITS7 and a 3:1 mix of the reverse primers ITS4 and ITS4arch, with both forward and reverse primers fitted with unique 8 bp sample identification tags (Ihrmark et al. 2012; Clemmensen et al. 2016). Composite amplicon samples were sequenced on the RSII platform (Pacific Biosciences) (Castaño et al. 2020) by SciLifeLab NGI (Uppsala, Sweden) after the addition of sequencing adaptors by ligation.

2.1 | Bioinformatic Analysis

For this study, OTU clustering was conducted using the SCATA pipeline (scata.mykopat.slu.se; Ihrmark et al. 2012) as described in Lindahl et al. (2021). An initial quality filtering removed sequences shorter than 100 bp with mean quality scores below 20, with individual bases quality scores under 3, with $< 90\%$ match to primer sequences, or lacking (or with mismatching) 3′ or 5′ tags. To reduce incidence of sequencing errors, globally unique genotypes were discarded. Quality filtering resulted in the removal of 53% of sequences, while an additional 16% were discarded as unique genotypes. The remaining sequences were clustered into SHs based on pairwise comparisons with USEARCH v11.0.667 (Edgar 2010), followed by single linkage clustering, in which a new sequence is added to a nearest OTU if it satisfies the matching condition for at least one sequence in that OTU. During the clustering, a minimum similarity to the closest neighbor required to enter a cluster was set to 99% (Kõljalg et al. 2013), and indels were penalized equally as substitutions.

ASVs were inferred using *DADA2* v 1.22.0 (Callahan et al. 2016) and the nf-core/ampliseq pipeline version 2.4.1 (Straub et al. 2020; Straub et al. 2023) based on the same filtered sequence data (i.e., after removal of low-quality reads and globally unique genotypes). The PacBio sequences were processed as single-end reads, and the *DADA2* package was used to denoise sequences and remove chimeras. Primer sequences were trimmed using Cutadapt v3.4 (Martin 2011), allowing for 10% mismatch, no insertion or deletions, and using the –pair-adapters function. Additional sequencing quality control was performed using FastQC v 0.11.9 (Andrews 2010). Co-occurrence of closely related ASVs was assessed, and potentially artifactual ASVs were filtered using the LULU package v 0.1.0 with default settings (Frøslev et al. 2017). LULU is a post-clustering curation method aiming at removing erroneous OTUs by merging of "daughter" OTUs with consistently co-occurring, similar and more abundant "parent" OTUs across a whole dataset, under the assumption that the "daughter" OTUs are artifacts in the sense that they represent genetic variation present in the species.

For ASVs, taxonomic labels were assigned directly to each exact sequence variant. For OTUs, annotations were based on the most abundant sequence within each cluster. In both cases, the sequences used for taxonomic assignment are hereafter referred to as "representative sequences." Taxonomic classification of representative sequences was based on the ITS2 region only and performed using nf-core/ampliseq v2.12 with Nextflow v2.10.3 (Di Tommaso et al. 2017) and the UNITE_alleuk database version 9.0 (Kõljalg et al. 2013; Abarenkov et al. 2023). Internally, ITSx version 1.1.3 (Bengtsson-Palme et al. 2013) is used to extract the ITS2 regions from the sequences. These sequences are then assigned to SHs by comparing them to the UNITE database using VSEARCH v2.21.1 (Rognes et al. 2016) and a sequence identity cut-off of 98.5% over the complete ITS2 region, and the taxonomic classification of the SH is adopted for the sequence. In case of no matches or matches to more than one SH, no SH is assigned and taxonomic classification is performed using the *assignTaxonomy* and *addSpecies* functions from the DADA2 package. These functions classify sequences using a naive Bayesian classifier against a reference database, assigning taxonomy from kingdom to genus (*assignTaxonomy*), followed by species-level annotation when exact matches are found (*addSpecies*). To assess within-OTU sequence heterogeneity, we randomly subsampled 200 sequences from the 10 most abundant OTUs and aligned them using MAFFT v7.526 with the --auto and --reorder options. We then calculated the average and minimum pairwise sequence similarity in Python using the *Bio.Phylo.TreeConstruction* module from Biopython, which computes uncorrected p-distances (proportion of differing nucleotide positions). Pairwise distance distributions were visualized as density plots.

2.2 | Statistical Analyses

All statistical analyses were conducted using R v.4.3.0 (R Core Team 2023), and data were visualized using *ggplot2* v3.5.2 (Wickham 2016), *VennDiagram* v1.7.3 (Chen 2022), *corrplot* 0.92

(Wei and Simko 2021), and *vegan* 2.6.4 (Oksanen et al. 2022). Alpha diversity was expressed as actual ASV/OTU richness (using the function specnumber), total extrapolated species richness (using function specpool), and as the Shannon diversity index (mean values ± SE) using the *vegan* package. Beta diversity was expressed as Jaccard and Bray–Curtis dissimilarity indexes (mean ± SE) using the function vegdist in *vegan*. The package *iNEXT* v3.0.0 was used to produce sample-size- and coverage-based rarefaction and extrapolation curves (Hsieh et al. 2016).

To assess sequence similarity between the two datasets, representative sequences were compared using VSEARCH v2.22.1 (Rognes et al. 2016). We used *--usearch_global* function to perform global pairwise alignment between OTU and ASV representative sequences. For this, one dataset (e.g., ASVs-based) was used as a reference database, and the other (e.g., OTU-based) as a query file. The same procedure was repeated in the reverse direction to ensure reciprocal matching. Alignments were filtered to include only matches with ≥90% sequence identity (*--id* 0.9) and at least 80% query coverage (*--query_cov* 0.80). Up to 50 best hits per query were retained (*--maxaccepts* 50), and all matches, including multiple hits per query, were recorded (*--uc_allhits*). When multiple hits occurred with identical identity and coverage, the top hit in the output order was retained for summary analyses.

Mantel tests were used to assess the Pearson's correlation between the ASV- and OTU-based Bray–Curtis dissimilarity matrices. The Bray–Curtis dissimilarity matrices were based on different taxonomic levels (e.g., phylum, class, order, family, genus, and species). Differences in median reads assigned to different taxonomic levels in ASV- and OTU-based datasets were compared using the Wilcoxon signed rank test with continuity correction.

Relationship between fungal community composition (both ASV- and OTU-based) and a selected environmental variable (soil N:C ratio; Lindahl et al. 2021) was tested using the adonis2 function in the *vegan* package. To account for potential temporal variation, we used restricted permutations (via the *strata* argument) with sampling year as a blocking factor.

3 | Results

3.1 | Outputs and Taxonomic Annotation of ASV and OTU Datasets

Clustering of metabarcoding reads using two different methods, resulting in OTUs and ASVs, yielded different numbers of detected units and affected the proportion of reads assigned to these units of diversity. Of the non-singleton, quality filtered reads in the OTU-based dataset, 11,214 reads (2.3%) were discarded in the process of generating ASVs. Nevertheless, 1343 (42%) more ASVs were generated compared to the number of OTUs (Table 1).

The percentage of reads assigned to different taxonomic ranks was similar across the ASV and OTU datasets; however, a slightly higher proportion of reads in the ASV dataset were classified at the genus level (78% of reads in ASVs, 73% of reads in ASVs after LULU curation, and 69% of reads in OTUs) and to the species level (63% of reads in ASVs, 57% of ASVs after LULU curation, and 46% of reads in OTUs) (Table 1). Further, the Mantel statistic indicated strong positive correlations between ASV and OTU datasets at all taxonomic levels (Table S1). The percentage of reads assigned to different taxonomic levels was consistently lower in the LULU curated ASV dataset compared to the initial ASV dataset. Further, fewer unique SH were identified after LULU curation. Together, this indicates that LULU curation frequently selects representative sequences that lack SH matches or higher taxonomic assignment.

Each of the two clustering methods determines a sequence to represent the units of diversity. All ASV representative sequences matched to OTU representative sequences with a minimum similarity of 92.2%. With the same minimum similarity, only 89% of OTU sequences matched an ASV sequence. The remaining 11% of OTUs were represented by sequences with very low similarity (< 92.2%) or no match at all among ASV sequences (Figure 1). However, all of these OTUs represent rare taxa, each comprising fewer than 15 reads and together accounting for less than 0.1% of the total reads clustered into OTUs (Figure 1a). It appears that reads captured by these rare OTUs were filtered out in the process of generating ASVs.

Multiple sequence alignment and pairwise distance calculation within the 10 largest OTUs using random subsampling of 200 sequences per OTU showed high internal similarity within all OTUs (average 97%–99.7%, minimum 92%–97%; Table S2; Figure S2), indicating that each OTU consists mainly of closely related sequence variants. However, in some of the investigated OTUs, we found clear internal structuring, representing distinct populations or cryptic species.

3.2 | Exploring Species Richness Estimates to Approximate Species Richness

We used the UNITE SH clustered at 1.5% dissimilarity across the ITS2 as a fair proxy of species of soil fungi. Generating OTUs with 99% single linkage clustering, like we did in this study, results in a dissimilarity between closely related OTUs comparable to UNITE SHs at 1.5%. In our OTU dataset, we captured 1766 unique SHs, together representing 1797 OTUs. In 28 cases, two OTUs matched the same SH, and in three cases, three OTUs matched the same SH. In the ASV dataset, on the other hand, we captured 1625 unique SHs represented by 2450 ASVs, with an average of 1.5 ASVs per unique SH. This indicates that ASVs capture both within and between species diversity, leading to splitting of taxa if ASVs are used as proxies for species. LULU curation did not resolve this issue, maintaining an ASV to SH ratio of 1.5.

Most SHs (1370) were captured by both datasets (Figure 2a). These accounted for 98.7% of the OTU reads assigned to SHs (Figure 2b). In our dataset, the 100 most abundant fungal taxa based on association with SHs (Supporting Information File S1) were consistently recovered using both ASV and OTU approaches, indicating robust identification of dominant species across methods. However, ASV clustering revealed additional fine-scale genetic variation within species that was masked in

TABLE 1 | Comparative analysis of ASV, ASV curated by LULU algorithm and OTU datasets outputs, including taxonomy, alpha, beta, and gamma diversity indices.

Parameter	ASV	ASV (LULU)	OTU
Overview			
Total reads	477,955	477,955	489,169
Observed richness (ASV/OTU)	4572	4208	3229
Number of UNITE SHs	1625	1534	1766
Reads assigned to SHs	253,271 (53%)	237,580 (50%)	307,568 (63%)
ASV/OTU assigned to SHs	2450 (54%)	2262 (54%)	1797 (56%)
Average reads in a single SH	103	105	171
Fungal taxonomy			
Assigned to phylum, % reads	97	91	96
Assigned to class, % reads	96	90	95
Assigned to order, % reads	95	89	94
Assigned to family, % reads	86	81	82
Assigned to genus, % reads	78	73	69
Assigned to species, % reads	63	57	46
Richness actual, average (±SE)	115 (±38)	113 (±37)	122 (±42)
Alpha-diversity index			
Shannon diversity, average (±SE)	3.6 (±0.5)	3.5 (±0.5)	3.3 (±0.6)
Beta-diversity index			
Jaccard dissimilarity, average (±SE)	0.937 (±0.045)	0.936 (±0.045)	0.891 (±0.067)
Bray–Curtis dissimilarity, average (±SE)	0.884 (±0.075)	0.883 (±0.076)	0.809 (±0.104)
Gamma-diversity index			
Richness (plot-based), extrapolated (±SE)	25,730 (±1832)	20,905 (±1493)	3535 (±32)

the OTU-based analysis. Further, 396 unique SHs were captured only by the OTU dataset and were not found in the ASV data; however, these were all rare, corresponding to 1.3% of all OTU reads that were assigned to SHs. In contrast, the 254 unique SHs that were only captured by the ASV dataset accounted for 14% of the ASV reads (Figure 2b). Among SHs that matched with both OTUs and ASVs, there was a strong positive relationship between ASV reads and OTU reads assigned to the same SHs ($p < 0.0001$; $R^2 = 0.99$, Figure S1), indicating coherent detection between datasets when compared at the level of SHs.

From the 100 most abundant soil fungi (Supporting Information File S1) we selected two commonly and broadly distributed species (*Piloderma sphaerosporum* SH0943263.09FU and *Piloderma olivaceum* SH0943302.09FU) to explore the species level resolution of the OTU and ASV as well as the effect of LULU curation of ASVs. Abundance and geographical distribution of the two species was similarly detected with both ASVs and OTUs (Figure 3). However, there were many more ASVs assigned to each of the selected *Piloderma* species compared to OTUs. Eight ASVs were identified as *P. sphaerosporum* and seven as *P. olivaceum* (Figure 3a,d), compared to only two and one OTUs assigned to the two species, respectively (Figure 3c,f). For both species, one or two ASVs represented the bulk of the read observations while the remaining ASVs were rare and detected only in a few plots. Both species were represented by a similar number of reads from a comparable number of plots in the two datasets (*P. sphaerosporum*: 12,337 reads in ASVs across 282 plots and 12,396 reads in OTUs across 252 plots; *P. olivaceum*: 1918 reads in ASVs across 90 plots and 1931 reads in OTUs across 71 plots). LULU curation did not merge ASVs sharing the same SH assignment for *P. sphaerosporum* and *P. olivaceum*, and the number of ASVs remained almost the same as before the curation (Figure 3b,e). Further, we found that the number of reads assigned to the two species after curation was similar, 12,658 and 1627 reads in ASV LULU for *P. sphaerosporum* and *P. olivaceum*, respectively (Figure 3).

3.3 | Species Richness Estimates

Irrespectively of how well the two methods serve as proxies for species, both the OTU and ASV data on average produced similar estimates of alfa species richness (122 OTUs and 115 ASVs per plot respectively, Table 1). Rarefaction analysis demonstrated that both methods reached saturation at the current

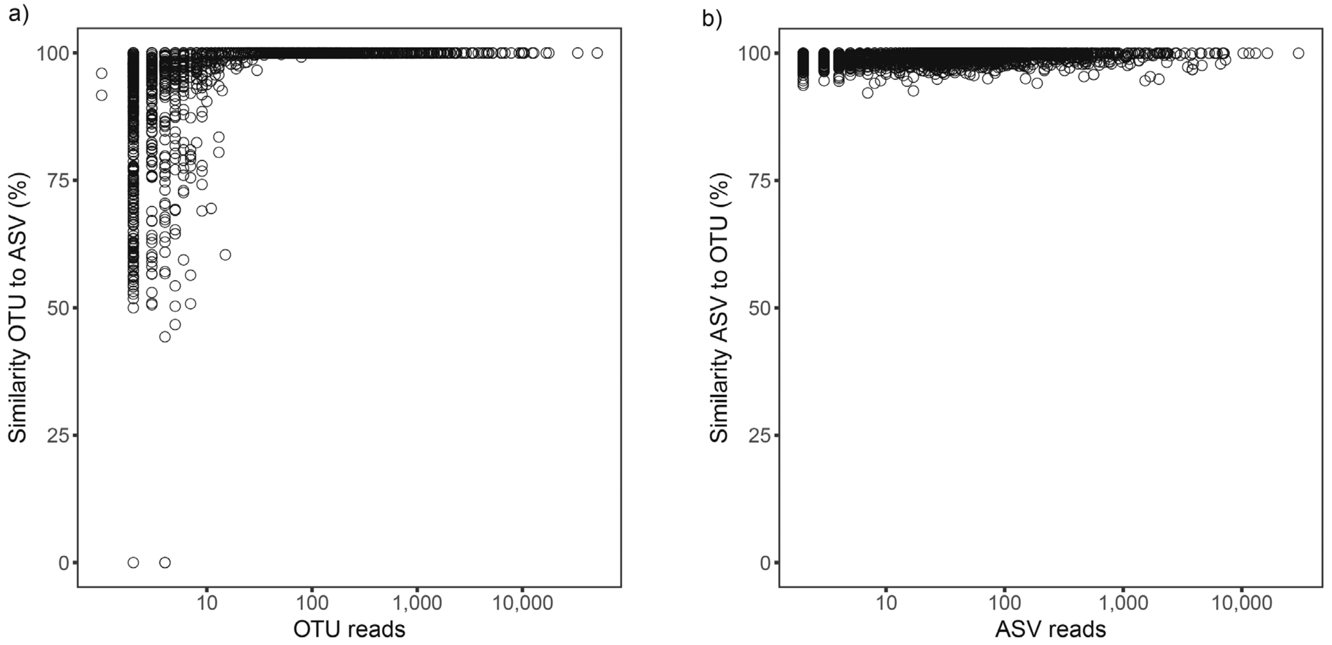

FIGURE 1 | Similarity of OTU sequences matched to ASV sequences (a) and ASV sequences matched to OTU sequences (b) in relation to their abundance on a log scale. Sequence similarity was established by matching the representative sequences from each dataset using global pairwise alignment implemented in VSEARCH v2.22.1.

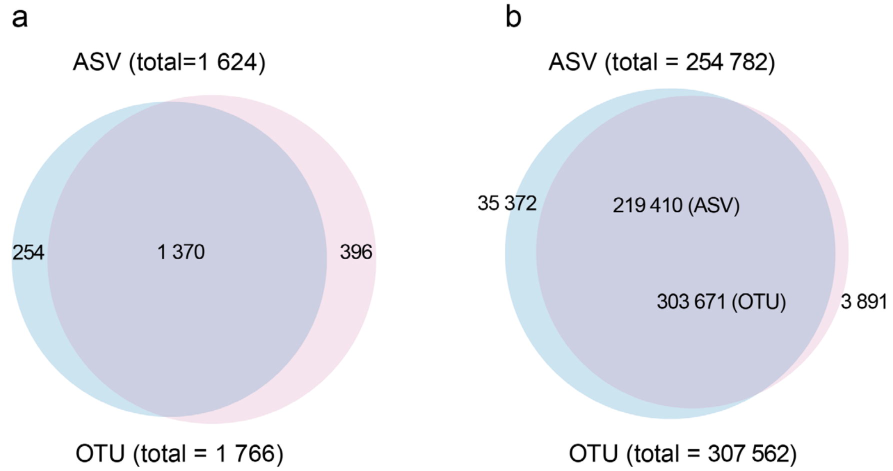

FIGURE 2 | The overlap and distribution of the number of UNITE Species Hypotheses (SHs) with matching ASVs and/or OTUs (a) and their corresponding read counts (b) across two datasets: ASVs (blue) and OTUs (pink). The proportions of the Venn diagram correspond to the relative sizes of the overlaps.

sequencing depth (Figure 4a) but at different richness estimates; the OTU dataset plateaued at around 3200, whereas the ASV dataset saturated at a total observed richness of just above 4500 ASVs (Table 1). Even with the 359 plots from conifer-dominating forest stands across Sweden, we found that detected species richness is expected to continue to increase with more plots included (Figure 4b). For the OTU dataset, the extrapolated richness reached an asymptote at around 3500—just some 300 taxa more than the observed richness. For the ASV dataset on the other hand, the total extrapolated richness was 6–7.3 times higher than the OTU-based estimate (Table 1). The splitting of taxa into multiple ASVs with somewhat different distribution, reflecting within species variance rather than between species variation (Figure 3), is likely a driver behind the much higher richness estimates.

3.4 | Correlation Between Different Diversity Metrics

We found that species richness estimates from ASVs and OTUs were strongly and positively correlated (Figure 5). Shannon diversity and evenness also correlated positively ($r = 0.95$). Fungal community composition derived from both ASV and OTU-based data correlated significantly with the selected environmental variable (soil N:C ratio). The PERMANOVA analysis indicated

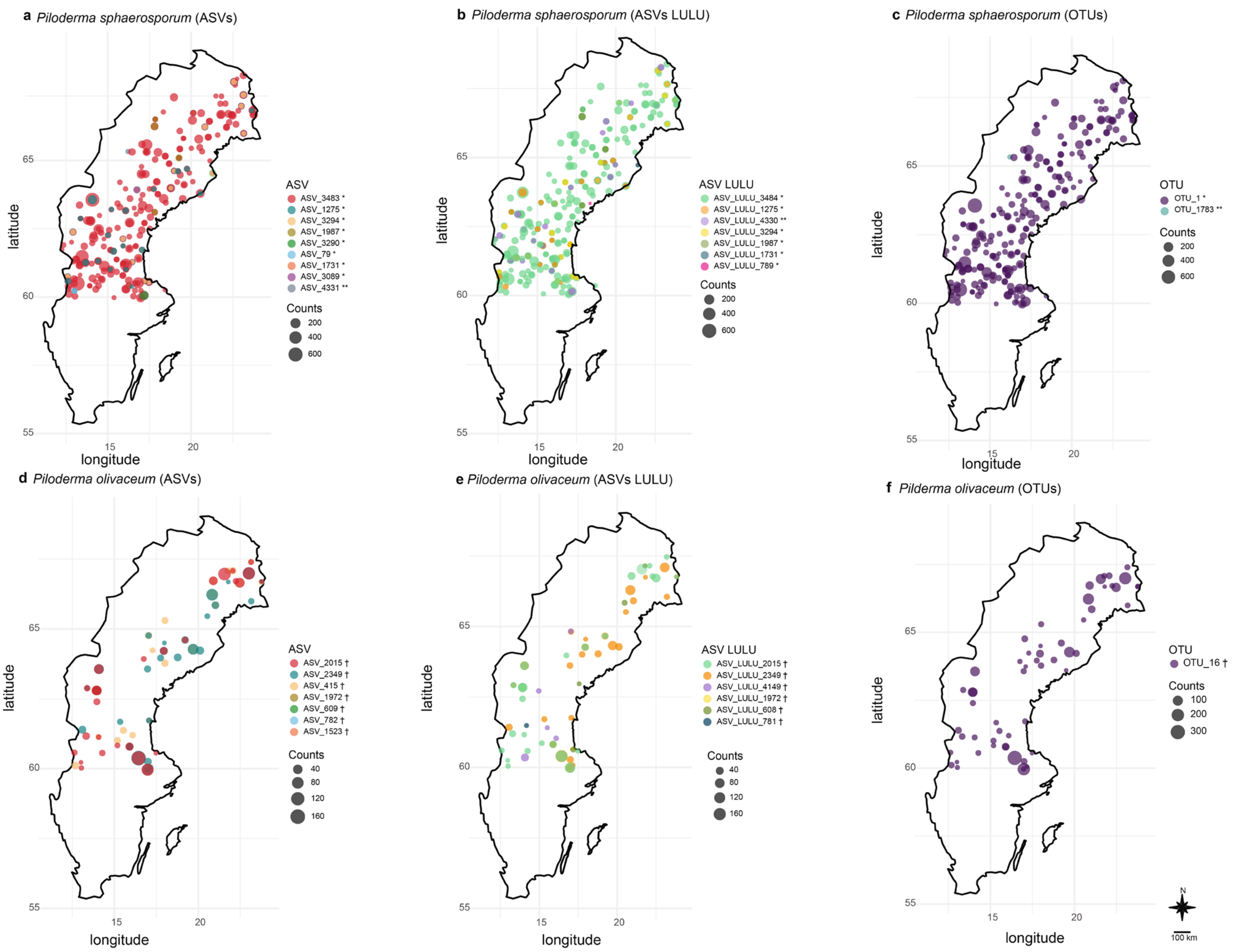

FIGURE 3 | Geographical distribution (indicated by longitude on the *x* axis and latitude on the *y* axis) of individual ASVs (a, d), ASVs LULU (b, e), and OTUs (c, f) assigned to *Piloderma sphaerosporum* (upper panel; *SH0943263.09FU; **SH0920530.09FU) and *Piloderma olivaceum* (lower panel; †SH0943302.09FU). The size of the circles corresponds to the abundance of each ASV/OTU, and colors indicate different ASVs/OTUs assigned to each *Piloderma* species. Only ASVs and OTUs that matched with UNITE Species Hypotheses (SHs) were included.

that N:C significantly correlated with the ASVs-based community composition ($F_{1357}=9.01$, $R^2=0.025$, $p=0.001$). Similarly, N:C had a significant correlation with OTU-based community composition ($F_{1,357}=14.82$, $R^2=0.040$, $p=0.001$), but N:C explained more of the variation in the OTU data. Similar levels of correlation were also observed when sampling year was included in the model as a random factor.

4 | Discussion

The comparison between ASV and OTU data indicated an overall high degree of overlap between the methods and highlighted the robustness of both approaches in their ability to reliably characterize soil fungi. However, the comparison also revealed some notable differences in the resolution and taxonomic assignments. In particular, 42% more ASVs compared to the number of OTUs were generated (30% more ASVs after LULU correction). Further, a slightly higher proportion of the ASVs were assigned (by *DADA2*) to genera and species compared to the OTUs.

Although ASVs and OTUs captured similar taxonomic structures of the fungal community, a number of OTUs did not match any of the ASVs, whereas all ASVs matched to OTUs. The presence of unmatched OTUs suggests that OTU clustering may retain rare or divergent sequence variants that are filtered out or lost during ASV denoising, likely because rare sequences are more sensitive to denoising thresholds and may be discarded as potential sequencing errors (Table 1; see also Joos et al. 2020; Tedersoo et al. 2022; Eshghi Sahraei et al. 2022). In contrast, the high degree of matching between ASVs and OTUs that could be assigned to UNITE SHs (Figure 2) further emphasizes the comparability of these methods. A majority of the OTU and ASV reads were assigned to identical UNITE SHs across both methods. However, a substantial proportion of the ASV reads were assigned to SHs that were not detected in the OTU dataset, indicating that ASVs may inflate species richness estimates by assigning unique sequence variants to additional taxa and SHs beyond those identified by OTUs (Figure 2), a pattern previously observed in several fungal metabarcoding studies (Callahan et al. 2017; Glassman and Martiny 2018; Tedersoo et al. 2022). Alternatively, OTUs might mask species richness by merging

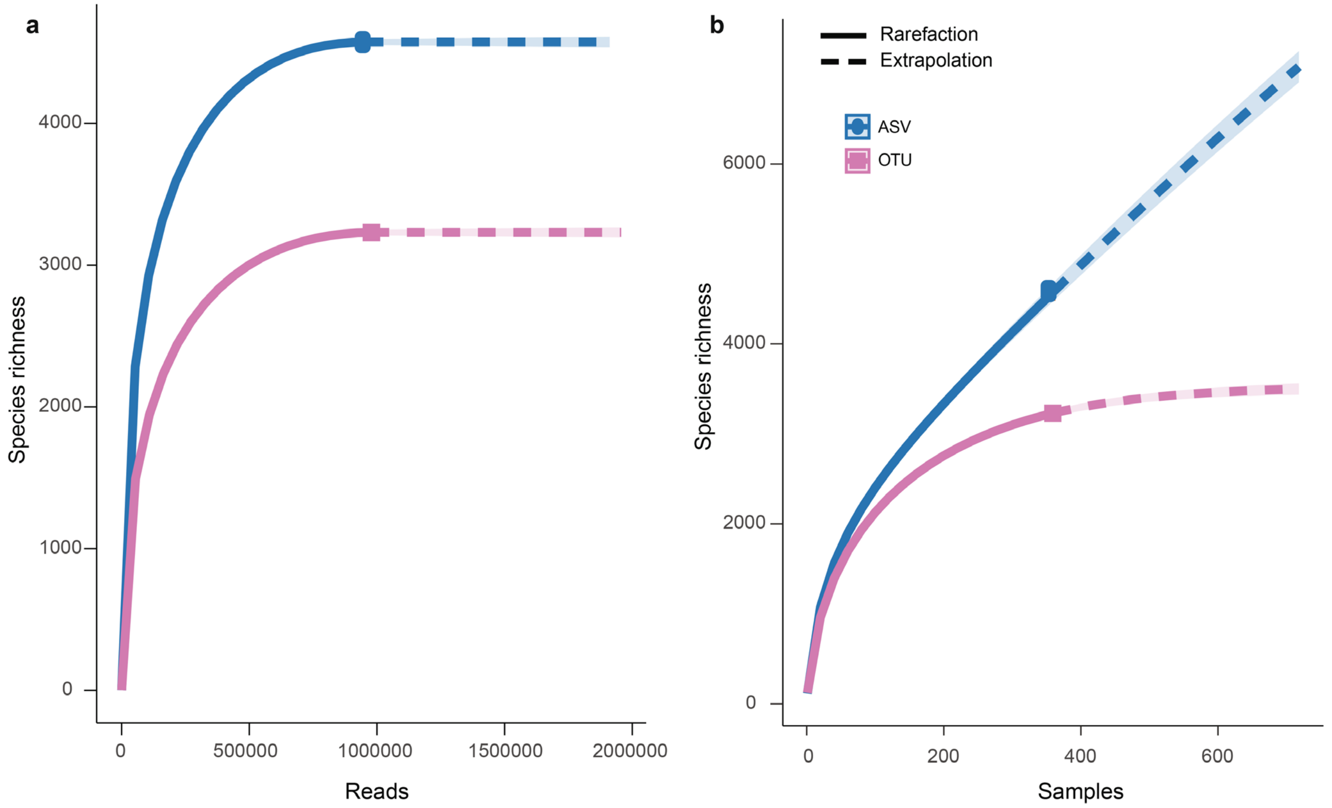

FIGURE 4 | Estimated species richness curves for rarefied samples (solid line) and extrapolated samples (dashed line) as a function of number of reads (a) and the number of samples (b) for ASVs (blue) and OTUs (purple) datasets. The 95% confidence intervals (light blue and light purple colors) were obtained by a bootstrap method based on 200 replications. The detected species richness is denoted by solid symbols. Each curve was extrapolated up to double its detected richness.

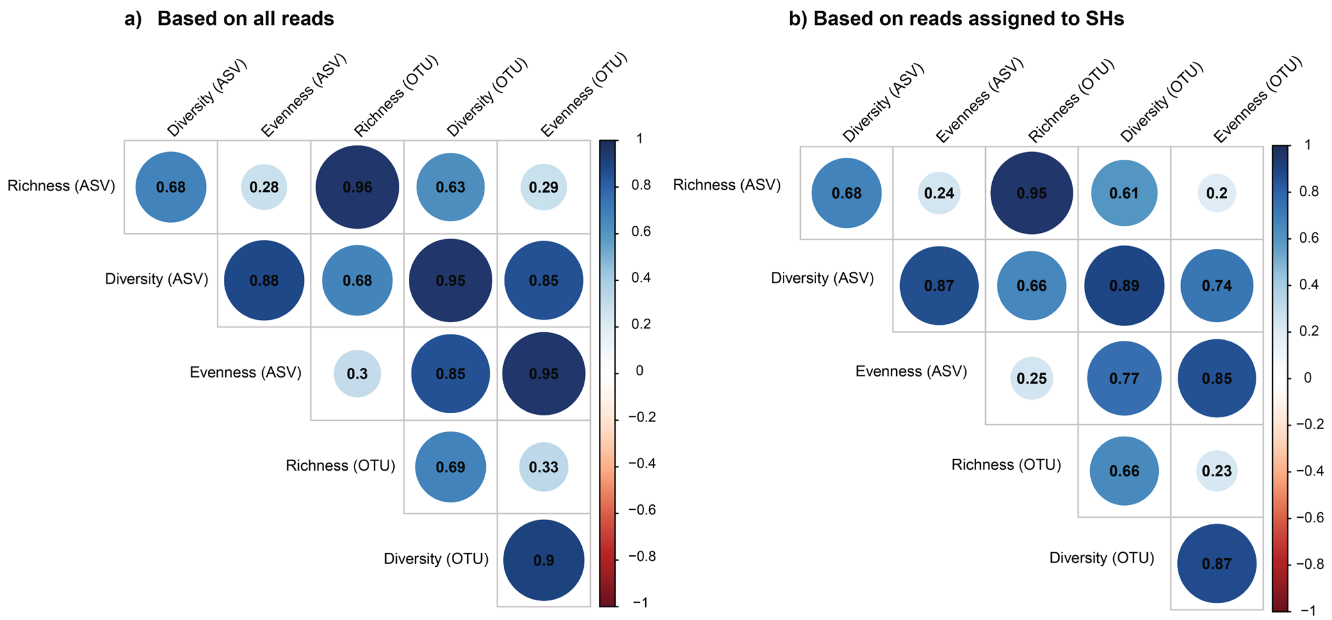

FIGURE 5 | Correlation matrix of Pearson's correlation matrix of ASV- and OTU-based alpha diversity metrics. Correlations between species richness, evenness, and Shannon diversity were calculated based on (a) all ASV and OTU read counts and (b) ASVs and OTUs read counts that were assigned to species hypotheses (SHs). Correlation coefficients are shown for each pairwise comparison if statistically significant ($p < 0.05$). The strength of the correlation is indicated by the color intensity—lighter shades represent weaker correlations, while darker blue tones indicate stronger ones.

closely related taxa. Further curation of reference databases is needed to refine SH assignments and better define the level of within-species variation. However, despite inconsistencies in how ASVs and OTUs assign to SHs, the most frequently observed variants/units in both datasets corresponded to the same SHs (Supporting Information File S1).

Extrapolated species richness was at least six times higher for ASVs compared to OTUs, reflecting the finer resolution and sensitivity of the ASV approach to intraspecific variation. We also found that the LULU algorithm resulted in a slightly lower number of ASVs (while keeping total reads the same); however, the curation did not resolve the overestimation of gamma species diversity by ASVs (Table 1). In contrast, OTU richness estimates reached an asymptote, indicating that estimation of species richness by the OTU method may require less sampling than estimation of genotype richness based on ASVs. ASVs identified more variants within common fungal species, such as *P. sphaerosporum* and *P. olivaceum*, compared to the OTU dataset. However, the overall abundance and geographical distribution of these species were similar across datasets, indicating that both methods reliably capture ecological patterns of dominant taxa. Therefore, ASVs should generally not be used as species proxies but need to be grouped post hoc to reflect diversity at the species level. Given the small proportion of fungal species that can be robustly identified to species level (e.g., Tedersoo et al. 2025), merging ASVs post hoc in this way is not feasible. At the same time, the ASV approach could be particularly valuable for capturing fine-scale genetic variation that is not captured in the OTU clustering in community analyses using ITS2 (Callahan et al. 2017). Many of the OTUs that we analyzed for internal variation contained two distinct subclusters, which may be ecologically informative and were obscured by OTU clustering.

Metabarcoding studies often focus on identifying taxa and characterizing community structure without necessarily resolving or emphasizing the taxonomic identities of the species involved. In such cases, the extent to which units/variants correspond to actual taxa/species is less critical, as long as richness and diversity estimates remain internally consistent and comparable within the study (Glassman and Martiny 2018). We found consistent alpha and beta diversity estimates between ASVs and OTUs (Figure 5; Table 1). Observed species richness (alpha diversity) was slightly higher for OTUs, with an average of 122 OTUs per plot compared to 115 ASVs. In contrast, beta-diversity indices were higher for ASV and ASV-LULU datasets than for OTUs, and gamma diversity estimates were six to seven times higher for ASVs, without yet reaching an asymptote (Figure 3b). This pattern—lower alpha but higher beta and gamma diversity for ASVs—is consistent with the idea that genotype-level resolution captures finer-scale variation, resulting in greater turnover across sites compared to OTU-based species-level clustering. The fact that ASV gamma diversity continues to increase with increasing sampling suggests that further sequencing would reveal additional intraspecific variation (Estensmo et al. 2021), indicating that a substantial portion of the regional genotype pool remained unsampled. In contrast, OTU-based gamma diversity appears to have plateaued. While ASVs can inflate diversity estimates at broader spatial scales, both ASV- and OTU-based approaches provide consistent insights into within-site species richness and community composition.

Soil N:C ratio was significantly correlated with community composition in both datasets, with a stronger correlation observed with the OTU-based data, in line with the higher ecological responsiveness of OTU data observed by Glassman and Martiny (2018). Given that ecological niches are expected to be less variable within biological species than between species, it seems logical that OTU clustering at the species level more efficiently captures ecological correlations than ASVs, which separate ecologically similar genotypes.

Overall, our findings indicate that despite differences in resolution and clustering methods, both ASV- and OTU-based approaches capture similar patterns in ITS-based soil fungal community composition. The strong correlations observed across multiple taxonomic levels support the reliability and validity of both methods for characterizing broad-scale fungal community patterns (Pauvert et al. 2019). ASVs dramatically inflate richness estimates, which cannot be accounted for with additional LULU-based curation. In contrast, the single-linkage OTUs used in this study offer a more robust estimate of soil fungal species richness. However, for studies focused on comparing richness patterns across treatments, either ASVs or OTUs may be appropriate. We argue that it is advantageous to use OTUs if the aim is to analyze ecological niches and community responses to environmental drivers, because response traits are homogenized by recombination within species, and splitting of species in many genotypes reduces correlation with environmental parameters. Ultimately, the choice between ASVs and OTUs should be guided by the specific research question and the desired level of taxonomic resolution.

Author Contributions

Conceptualization: J.K., B.D.L., and A.R. Bioinformatics: J.K., J.T., and B.D.L. Data analyses: J.K. and A.R. Visualization: J.K. Writing – original draft: J.K., A.R. Writing – review and editing: J.K., J.T., A.D., B.D.L., and A.R. Supervision: A.R. Funding acquisition: A.R.

Acknowledgments

This research was supported by Swedish Biodiversity Data Infrastructure (SBDI).

Funding

This work was supported by Swedish Biodiversity Data Infrastructure funded by VR 2019-00242.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in SRA at https://www.ncbi.nlm.nih.gov/sra, reference number PRJNA693127.

References

Abarenkov, K., R. H. Nilsson, K.-H. Larsson, et al. 2024. "The UNITE Database for Molecular Identification and Taxonomic Communication of Fungi and Other Eukaryotes: Sequences, Taxa and Classifications Reconsidered." *Nucleic Acids Research* 52: 791–797.

Abarenkov, K., A. Zirk, T. Piirmann, et al. 2023. *Full UNITE+INSD Dataset for Eukaryotes. Version 18.07.2023*. UNITE Community.

Andrews, S. 2010. "FastQC a Quality Control Tool for High Throughput Sequence Data [Online]." http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bengtsson-Palme, J., M. Ryberg, M. Hartmann, 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.

Blaxter, M., J. Mann, T. Chapman, et al. 2005. "Defining Operational Taxonomic Units Using DNA Barcode Data." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 360: 1935–1943.

Callahan, B., P. McMurdie, and S. Holmes. 2017. "Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis." *ISME Journal* 11: 2639–2643.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. "DADA2: High-Resolution Sample Inference From Illumina Amplicon Data." *Nature Methods* 13: 581–583.

Castaño, C., A. Berlin, M. Brandström Durling, et al. 2020. "Optimized Metabarcoding With Pacific Biosciences Enables Semi-Quantitative Analysis of Fungal Communities." *New Phytologist* 228: 1149–1158.

Chen, H. 2022. "VennDiagram: Generate High-Resolution Venn and Euler Plots." R package version 1.7.3.

Chiarello, M., M. McCauley, S. Villéger, and C. R. Jackson. 2022. "Ranking the Biases: The Choice of OTUs vs. ASVs in 16S rRNA Amplicon Data Analysis Has Stronger Effects on Diversity Measures Than Rarefaction and OTU Identity Threshold." *PLoS One* 17: e0264443.

Clemmensen, K. E., K. Ihrmark, M. Brandström Durling, and B. D. Lindahl. 2016. "Sample Preparation for Fungal Community Analysis by High-Throughput Sequencing of Barcode Amplicons." In *Microbial Environmental Genomics*, edited by F. Martin and S. Uroz, 55–88. Springer.

Di Tommaso, P., M. Chatzou, E. W. Floden, P. P. Barja, E. Palumbo, and C. Notredame. 2017. "Nextflow Enables Reproducible Computational Workflows." *Nature Biotechnology* 35: 316–319.

Edgar, R. C. 2010. "Search and Clustering Orders of Magnitude Faster Than BLAST." *Bioinformatics* 26: 2460–2461.

Edgar, R. C. 2013. "UPARSE: Highly Accurate OTU Sequences From Microbial Amplicon Reads." *Nature Methods* 10: 996–998.

Egan, C. P., A. Rummel, V. Kokkoris, J. Klironomos, Y. Lekberg, and M. Hart. 2018. "Using Mock Communities of Arbuscular Mycorrhizal Fungi to Evaluate Fidelity Associated With Illumina Sequencing." *Fungal Ecology* 33: 52–64.

Eren, A. M., L. Maignien, W. J. Sul, et al. 2013. "Oligotyping: Differentiating Between Closely Related Microbial Taxa Using 16S rRNA Gene Data." *Methods in Ecology and Evolution* 4: 1111–1119.

Eshghi Sahraei, S., B. Furneaux, K. Kluting, et al. 2022. "Effects of Operational Taxonomic Unit Inference Methods on Soil Microeukaryote Community Analysis Using Long-Read Metabarcoding." *Ecology and Evolution* 12: e8676.

Estensmo, E. L. F., S. Maurice, L. Morgado, P. M. Martin-Sanchez, I. Skrede, and H. Kauserud. 2021. "The Influence of Intraspecific Sequence Variation During DNA Metabarcoding: A Case Study of Eleven Fungal Species." *Molecular Ecology Resources* 21: 1141–1148.

Fridman, J., S. Holm, M. Nilsson, P. Nilsson, A. Hedström Ringvall, and G. Ståhl. 2014. "Adapting National Forest Inventories to Changing Requirements —The Case of the Swedish National Forest Inventory at the Turn of the 20th Century." *Silva Fennica* 48: 1095.

Frøslev, T. G., R. Kjøller, H. H. Bruun, et al. 2017. "Algorithm for Post-Clustering Curation of DNA Amplicon Data Yields Reliable Biodiversity Estimates." *Nature Communications* 8: 1188.

Furneaux, B., M. Bahram, A. Rosling, N. S. Yorou, and M. Ryberg. 2021. "Long- and Short-Read Metabarcoding Technologies Reveal Similar Spatiotemporal Structures in Fungal Communities." *Molecular Ecology Resources* 21: 1833–1849.

Glassman, S. I., and J. B. H. Martiny. 2018. "Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants Versus Operational Taxonomic Units." *mSphere* 3: e00148-18. https://doi.org/10.1128/msphere.00148-18.

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

Huse, S. M., D. M. Welch, H. G. Morrison, and M. L. Sogin. 2010. "Ironing Out the Wrinkles in the Rare Biosphere Through Improved OTU Clustering." *Environmental Microbiology* 12: 1889–1898.

Ihrmark, K., I. T. M. Bödeker, K. Cruz-Martinez, et al. 2012. "New Primers to Amplify the Fungal ITS2 Region—Evaluation by 454-Sequencing of Artificial and Natural Communities." *FEMS Microbiology Ecology* 82: 666–677.

Joos, L., S. Beirinckx, A. Haegeman, et al. 2020. "Daring to be Differential: Metabarcoding Analysis of Soil and Plant-Related Microbial Communities Using Amplicon Sequence Variants and Operational Taxonomical Units." *BMC Genomics* 21: 1–17.

Kauserud, H. 2023. "ITS Alchemy: On the Use of ITS as a DNA Marker in Fungal Ecology." *Fungal Ecology* 65: 101274.

Kõljalg, U., K.-H. Larsson, K. Abarenkov, et al. 2005. "UNITE: A Database Providing Web-Based Methods for the Molecular Identification of Ectomycorrhizal Fungi." *New Phytologist* 166: 1063–1068.

Kõljalg, U., H. R. Nilsson, D. Schigel, et al. 2020. "The Taxon Hypothesis Paradigm—On the Unambiguous Detection and Communication of Taxa." *Microorganisms* 8: 1910.

Kõljalg, U., R. H. Nilsson, K. Abarenkov, et al. 2013. "Towards a Unified Paradigm for Sequence-Based Identification of Fungi." *Molecular Ecology* 22: 5271–5277.

Krah, F.-S., and M. March-Salas. 2022. "eDNA Metabarcoding Reveals High Soil Fungal Diversity and Variation in Community Composition Among Spanish Cliffs." *Ecology and Evolution* 12: e9594.

Lindahl, B. D., J. Kyaschenko, K. Varenius, et al. 2021. "A Group of Ectomycorrhizal Fungi Restricts Organic Matter Accumulation in Boreal Forest." *Ecology Letters* 24: 1341–1351.

Lindahl, B. D., R. H. Nilsson, L. Tedersoo, et al. 2013. "Fungal Community Analysis by High-Throughput Sequencing of Amplified Markers—A User's Guide." *New Phytologist* 199: 288–299.

Lindner, D. L., T. Carlsen, R. H. Nilsson, M. Davey, T. Schumacher, and H. Kauserud. 2013. "Employing 454 Amplicon Pyrosequencing to Reveal Intragenomic Divergence in the Internal Transcribed Spacer rDNA Region in Fungi." *Ecology and Evolution* 3: 1751–1764.

Lücking, R., M. C. Aime, B. Robbertse, et al. 2020. "Unambiguous Identification of Fungi: Where Do We Stand and How Accurate and Precise Is Fungal DNA Barcoding?" *IMA Fungus* 11: 14.

Maeda, T., Y. Kobayashi, H. Kameoka, et al. 2018. "Evidence of Non-Tandemly Repeated rDNAs and Their Intragenomic Heterogeneity in *Rhizophagus irregularis*." *Communications Biology* 1: 87.

Martin, M. 2011. "Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads." *EMBnet.Journal* 17: 10–12.

Nguyen, N. H., Z. Song, S. T. Bates, et al. 2016. "FUNGuild: An Open Annotation Tool for Parsing Fungal Community Datasets by Ecological Guild." *Fungal Ecology* 20: 241–248.

Nilsson, R. H., S. Anslan, M. Bahram, C. Wurzbacher, P. Baldrian, and L. Tedersoo. 2019. “Mycobiome Diversity: High-Throughput Sequencing and Identification of Fungi.” *Nature Reviews Microbiology* 17: 95–109.

Nilsson, R. H., K.-H. Larsson, A. F. S. Taylor, et al. 2019. “The UNITE Database for Molecular Identification of Fungi: Handling Dark Taxa and Parallel Taxonomic Classifications.” *Nucleic Acids Research* 47: D259–D264.

Oksanen, J., G. Simpson, F. Blanchet, et al. 2022. “Vegan: Community Ecology Package.” R package version 2.6–4.

Pauvert, C., M. Buée, V. Laval, et al. 2019. “Bioinformatics Matters: The Accuracy of Plant and Soil Fungal Community Data Is Highly Dependent on the Metabarcoding Pipeline.” *Fungal Ecology* 41: 23–33.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Renaudin, M., R. Khlifa, S. Legault, et al. 2023. “Long-Term Simulated Nitrogen Deposition Has Moderate Impacts on Soil Microbial Communities Across Three Bioclimatic Domains of the Eastern Canadian Forest.” *Forests* 14: 1124.

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

Runnel, K., P. Lõhmus, K. Küngas, et al. 2024. “Aerial eDNA Contributes Vital Information for Fungal Biodiversity Assessment.” *Journal of Applied Ecology* 61: 2418–2429.

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

Straub, D., N. Blackwell, A. Langarica-Fuentes, A. Peltzer, S. Nahnsen, and S. Kleindienst. 2020. “Interpretations of Environmental Microbial Community Studies Are Biased by the Selected 16S rRNA (Gene) Amplicon Sequencing Pipeline.” *Frontiers in Microbiology* 11: 550420.

Straub, D., J. Tångrot, A. Peltzer, et al. 2023. “Nf-core/Ampliseq: Ampliseq Version 2.12.0 [Software].” *Zenodo*. https://doi.org/10.5281/zenodo.1493841.

Tedersoo, L., M. Bahram, S. Põlme, et al. 2014. “Fungal Biogeography. Global Diversity and Geography of Soil Fungi.” *Science* 346: 1256688.

Tedersoo, L., M. S. Hosseyni Moghadam, K. Panksep, et al. 2025. “Thirty Novel Fungal Lineages: Formal Description Based on Environmental Samples and DNA.” *MycoKeys* 124: 1–121.

Tedersoo, L., V. Mikryukov, A. Zizka, et al. 2022. “Global Patterns in Endemicity and Vulnerability of Soil Fungi.” *Global Change Biology* 28: 6696–6710.

Wei, T., and V. Simko. 2021. “R Package ‘Corrplot’: Visualization of a Correlation Matrix.” Version 0.92.

Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Zhenting, S., Y. Limin, Y. Meiling, L. Kexin, Y. Li, and H. Mei. 2024. “Temporal Heterogeneity of the Root Microbiome in *Panax ginseng* Soils Across Ecological Compartments Under Mild Soil Disturbance.” *Frontiers in Microbiology* 15: 1340575.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Comparison of the top 100 most abundant ASVs and corresponding OTUs based on read count and frequency. **Figure S1:** Relationship between the number of reads (log scale) assigned to Species Hypotheses (SHs) in ASV- and OTU-based datasets. Each point on the scatterplot represents a SH that is shared by both datasets. **Figure S2:** Density plots of pairwise sequence differences within the 10 largest SCATA OTUs (200 randomly subsampled sequences per OTU). Pairwise differences were calculated as percent sequence divergence, and the *x*-axis is truncated to 90%–100% to highlight high-divergence patterns. **Table S1:** Mantel test results. **Table S2:** Number of sequences, average and minimum similarity within the 10 largest OTUs. The similarity was calculated using random subsampling of 200 sequences per OTU.