EukFunc: A Holistic Eukaryotic Functional Reference for Automated Profiling of Soil Eukaryotes

Lentendu, Guillaume; Singer, David; Agatha, Sabine; Bahram, Mohammad; Hannula, S.E. et al

RESOURCE ARTICLE

EukFunc: A Holistic Eukaryotic Functional Reference for Automated Profiling of Soil Eukaryotes

Guillaume Lentendu[1] | David Singer[2] | Sabine Agatha[3] | Mohammad Bahram[4,5,6] | S. Emilia Hannula[7] | Johannes Helder[8] | Leho Tedersoo[9,10] | Walter Traunspurger[11] | Stefan Geisen[8] | Enrique Lara[12]

[1]Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland | [2]Changins College for Viticulture and Enology, University of Applied Sciences and Art Western Switzerland, Nyon, Switzerland | [3]Department of Environment & Biodiversity, Paris Lodron University of Salzburg, Salzburg, Austria | [4]Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden | [5]Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia | [6]Department of Agroecology, Aarhus University, Slagelse, Denmark | [7]Department of Environmental Biology, Institute of Environmental Sciences (CML), Leiden University, Einsteinweg 2, Leiden, the Netherlands | [8]Laboratory of Nematology, Wageningen University, Wageningen, the Netherlands | [9]Mycology and Microbiology Center, University of Tartu, Tartu, Estonia | [10]College of Science, King Saud University, Riyadh, Saudi Arabia | [11]Animal Ecology, Bielefeld University, Bielefeld, Germany | [12]Real Jardín Botánico de Madrid—CSIC, Madrid, Spain

Correspondence: Guillaume Lentendu (guillaume.lentendu@unine.ch)

ABSTRACT

The soil eukaryome constitutes a significant portion of Earth's biodiversity that drives major ecosystem functions, such as controlling carbon fluxes and plant performance. Currently, however, we miss a standardised approach to functionally classify the soil eukaryome in a holistic way. Here we compiled EukFunc, the first functional reference database that characterises the most abundant and functionally important soil eukaryotic groups: fungi, nematodes and protists. We classified the 14,060 species in the database based on their mode of nutrient acquisition into the main functional classes of symbiotroph (40%), saprotroph (26%), phototroph (17%), predator (16%) and unknown (2%). EukFunc provides further detailed information about nutrition mode, including a secondary functional class (i.e., for organisms with multiple nutrition modes), and preyed or associated organisms for predatory or symbiotic taxa, respectively. EukFunc is available in multiple formats for user-friendly functional analyses of specific taxa or annotations of metabarcoding datasets, both embedded in the R package EukFunc. Using a soil dataset from alpine and subalpine meadows, we highlighted the extended ecological insights obtained from combining functional information across the entire soil eukaryome as compared to focusing on fungi, protists or nematodes individually. EukFunc streamlines the annotation process, enhances efficiency and accuracy, and facilitates the investigation of the functional roles of soil eukaryotes—a prerequisite to better understanding soil systems.

Stefan Geisen and Enrique Lara contributed equally to the paper.

1 | Introduction

The soil eukaryome encompasses diverse organisms, including fungi, metazoans and protists (Baldrian et al. 2022; Oliverio et al. 2020). These organisms contain millions of species and represent much of Earth's biodiversity (Anthony et al. 2023). Currently, virtually all soil biodiversity studies taxonomically characterise distinct parts of this eukaryotic biodiversity using metabarcoding, mostly targeting the ITS region for fungi (Baldrian et al. 2022) and the 18S rRNA genes for metazoans and protists (Lentendu et al. 2023). These metabarcoding approaches provide valuable information about the composition of distinct soil eukaryotic communities (e.g., being restricted to fungi), often applying taxon-specific genomic regions as targets, whereas a comprehensive overview of all groups of eukaryotes remains scarce (e.g., combining fungal sequencing of the ITS region and protist sequencing of the 18S rRNA gene). Generally, these metabarcoding-based taxonomic community profiling approaches do not directly provide functional information on the detected taxa and communities and must subsequently be translated to functional traits based on our knowledge of species ecology. Among these traits are those related to nutrient acquisition that are shared across the whole Domain Eukarya.

All eukaryotes can functionally be broadly classified into phototrophs (e.g., photosynthetic organisms), predators (organisms that engulf others), saprotrophs (organisms that use dead organic matter as the nutrient source through decomposing it) and symbiotrophs (organisms linked to heterospecific organisms in any association). These classes are sometimes combined; for instance, facultative pathogens, like many *Fusarium* species (Fungi) and Peronosporomycetes (=Oomycetes; protists) are both symbiotrophs and saprotrophs. These broad categories can be further divided into sub-categories. For instance, symbiotrophs might be mutualists or parasites. Thus, the consideration of multiple functions in the numerically dominant groups of soil eukaryotes contributes to a more comprehensive characterisation of the functional diversity. Eventually, these individual eukaryotic functions synergistically determine the soil's and ecosystem's functions, including nutrient cycling, carbon storage and primary production (Geisen et al. 2018; Wagg et al. 2014).

Assigning more detailed functional diversity of eukaryotes using sequence-based assignments might even be straightforward, as the limited occurrence of horizontal gene transfer between taxa ensures functional conservation within clades (Lara et al. 2022). Accordingly, tools for functional placements of specific groups of soil eukaryotes exist. For fungi, sequence data can be processed with FUNGuild (Nguyen et al. 2016) or FungalTraits (Põlme et al. 2020), which automatically classifies fungal genera into main functional types. Soil nematode sequences can be functionally annotated using the web interface Ninja (Sieriebriennikov et al. 2014). For protists (Geisen et al. 2018), however, no automated tool for functional annotations is available; hence, manual processing is inevitable (e.g., Dumack et al. 2020). The different metabarcoding approaches, group-specific functional annotation tools, as well as limited scientist-specific expertise, often determine the study focus and prevent a holistic functional annotation of the soil eukaryome.

To overcome these shortcomings, we present the first functionally annotated database of the soil eukaryome, including fungi, nematodes and protists, arguably the most abundant and diverse groups in soil that thereby dominate metabarcoding studies. This database supports metabarcoding approaches and is built upon the established PR^2 reference database (Guillou et al. 2013). Finally, we exemplified the additional value of the R package *EukFunc* by the analyses of a metabarcoding 18S rRNA gene (V4 variable domain) dataset from 161 soil samples collected from alpine and subalpine meadows (Mazel et al. 2021; Seppey et al. 2020).

2 | Materials and Methods

2.1 | Terminology

To provide a consistent functional assignment across the whole eukaryome, the function of an organism has been defined according to the main mode of acquisition of nutrients and energy. Predators are those that eat other organisms (note, for predators we use a broad ecological concept of actively preying on other organisms; this definition contrasts with the classical finer groupings like for nematodes that distinguish bacterivores, fungivores, omnivores and animal predators). Symbiotrophs live in association with other organisms, including parasitism, mutualism or anything in between. Saprotrophs take both energy and nutrients from dead organisms or organic matter. Phototrophs are organisms that perform photosynthesis, regardless of whether they perform other functions (e.g., predation, symbiotrophy) simultaneously. Combinations of these main categories apply sometimes for organisms with two types of energy acquisition; for instance, the mixotrophic soil dinoflagellate *Gloeodinium montanum* is both a phototroph and a predator (Patterson and Larsen 1992). Besides the general assignment to the main categories, further detailed information regarding the organism's function is provided (see below).

2.2 | Database Content and Sources

The database developed for this study is based on the nine ranks of taxonomic nomenclature of the PR^2 database v5.0.0 (Guillou et al. 2013) and provides functional assignments for terrestrial fungi, nematodes and protists. For fungi, FunGuild was the main resource for functional annotations (Nguyen et al. 2016). Nematodes were functionally placed using an internal database, originally provided in the NemaGuild online database and manually curated (http://stbates.org/nemaguild_db.php). The functional annotations for protists were primarily derived from Berger (1999), Foissner (1993), Lynn (2008) and multiple other references for ciliates (Alveolata, Ciliophora; listed in the 'reference' column of the DBu and DBf databases), and on a comprehensive literature search mostly based on Adl et al. (2019) for all other protists. To ensure quality control, comparisons with other databases and sources, such as those of Faure et al. (2019) and Dumack et al. (2020), were performed. Based on this literature,

exclusively aquatic clades available in PR² were removed from the database.

The R package *EukFunc* grants access to the database, of which six different versions are available (Table S1): (i) *DBf* includes all PR² accessions of terrestrial protists, fungi and nematodes; (ii) *DBu* includes each unique species present in the database; (iii) *DBc* includes each clade for which a single functional assignment is available across all fields; (iv) *DBu_main* includes each unique species while expanding the class of symbiotrophs for parasites and host phototrophs; (v) *DBc_main* includes each clade for which a single main functional class can be assigned, while expanding the class of symbiotrophs for parasites and host phototrophs (i. e. symbiotrophic organism hosting one or multiple phototrophic organisms); and (vi) *DBU_minimal* includes each clade for which a single main functional class can be assigned.

2.3 | Functional Database of Soil Eukaryotes

The functional database presented in this study provides unique information regarding the trophic mode of each species, as well as associated organism(s) and the environment(s) the species inhabit. Up to two main functional classes are given (Figure 1, Table 1). They comprise predator, symbiotroph, phototroph, saprotroph (see definition above) and unknown. The imprecise term 'mixotroph' is not used in the database as it simply refers to unicellular organisms capable of obtaining energy from two different sources. Mixotrophic protists can however, be inferred based on their main and secondary functional classes. For soil protists, they fall into two broad categories (1) constitutive mixotrophs that perform phagocytosis and use their plastids to synthesise nutrients (like in genus *Euglena*) or (2) predators and symbiotrophs that obtain nutrients and energy both through phagocytosis and through the photosynthesis products of their hosted photobionts (e.g., *Hyalosphenia papilio*, an arcellinid testate amoeba which predate bacteria and smaller protists while hosting phototrophic algae; see

Table S2); in that case, the secondary functional class mentions 'host phototroph'. This approach was chosen purposefully as some mixotrophs are predominantly phototrophs or predators, and the ranking of the main feeding type allows for a differentiation.

Frequently, details about the trophic status are provided in separate fields such as 'detailed_functional_class', 'detailed_secondary_functional_class' (e.g., 'algae and protist parasite', 'litter saprotroph; soil saprotroph', 'bacteria and small protist phagotroph') (Table S3). For fungi, these data correspond to the guilds in the FunGuild database. The 'assignment_level' indicates the taxonomic rank at which each functional assignment is shared among clades. A confidence level is provided for the functional assignment of fungi, ranging from 'possible' via 'probable' to 'highly probable'. This may be extended to all clades in future database versions when knowledge increases.

The fields 'associated organism', 'associated material' and 'environment' are primarily derived from the detailed functional class fields and provide convenient selection vectors when specific organisms, materials or environments are interesting. The field 'potential human pathogens' marks organisms that are known to be found in soil, mostly as dormant life stages (Santamaría and Toranzos 2003); they are designated as 'opportunistic' or 'probably opportunistic' (Table S4). Finally, a generic 'comment' field is provided to offer further precision when necessary, and a 'reference' field specifies the primary source of information for the functional assignment, including taxonomic expert names or laboratories, publication DOIs or former database names. Multiple independent information can be provided in each field, separated by semicolons.

2.4 | R Package Usage and Function

The R package *EukFunc* has been developed not only to provide the functional database but also to include helper functions for linking
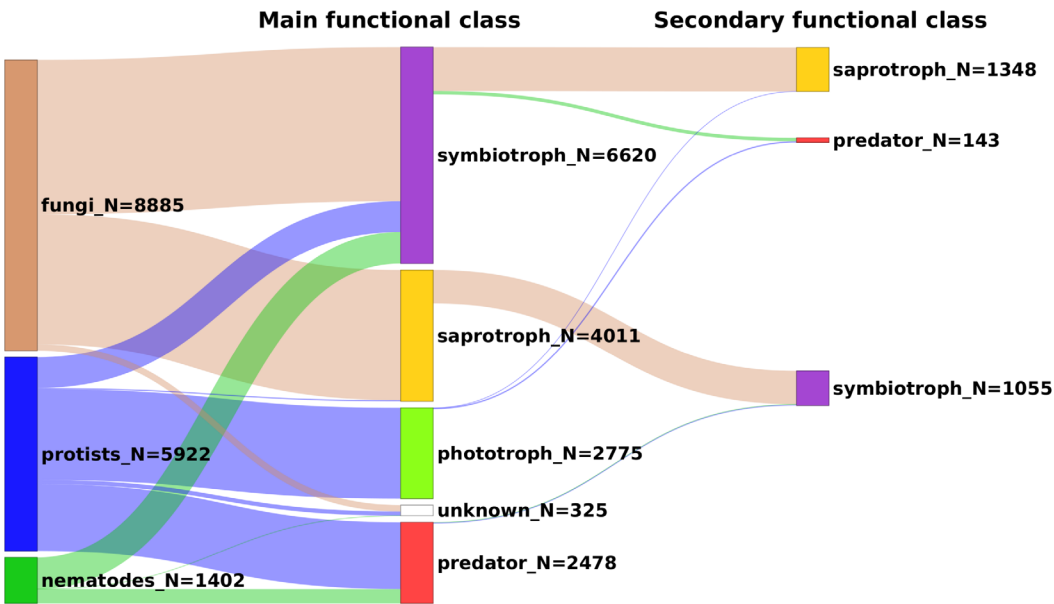

FIGURE 1 | Distribution of the main functional and secondary functional classes of the EukFunc database across species of terrestrial fungi, nematodes and protists considered from the PR² database.

TABLE 1 | Main and secondary functional assignments of species contained in the database.

Clade	Main functional class	Secondary functional class	Species no (%)	Sequences no (%)
Fungi	saprotroph		2959 (33.3)[a]	10,128 (18.4)
	saprotroph	& symbiotroph: host phototroph	10 (0.1)	17 (0)
		& symbiotroph: others	581 (6.5)	1713 (3.1)
		& symbiotroph: parasite	1773 (20)	8647 (15.7)
	symbiotroph: host phototroph		692 (7.8)	1610 (2.9)
	symbiotroph: others		746 (8.4)	4049 (7.4)
	symbiotroph: parasite		1930 (21.7)	11,586 (21.1)
	unknown		194 (2.2)	17,252 (31.4)
Nematodes	predator		425 (30.3)	1443 (32.3)
	predator	& symbiotroph: others	2 (0.1)	2 (0)
	predator	& symbiotroph: parasite	113 (8.1)	350 (7.8)
	symbiotroph: parasite		860 (61.3)	2661 (59.6)
	unknown		2 (0.1)	11 (0.2)
Protists	phototroph		2734 (46.2)	14,132 (32.2)
	phototroph	& predator	38 (0.6)	180 (0.4)
		& saprotroph	3 (0.1)	12 (0)
	predator		2017 (34.1)	15,253 (34.8)
	predator	& symbiotroph: host phototroph	21 (0.3)	980 (2.1)
		& symbiotroph: others	4 (0.1)	33 (0.1)
		& symbiotroph: parasite	1 (0)	4 (0)
	saprotroph		33 (0.6)	174 (0.4)
	symbiotroph: host phototroph		33 (0.6)	174 (0.4)
	symbiotroph: others		189 (3.2)	814 (1.9)
	symbiotroph: parasite		753 (12.7)	8834 (20.1)
	unknown		129 (2.2)	3447 (7.9)

[a]Clades' computed percentage in parentheses.

functional information to taxonomy tables generated through standard bioinformatic analyses (*sensu* Lentendu et al. 2023). The main function of the package is 'assign_path', which accepts a taxonomy data frame or vector as input and performs a partial match of the provided taxonomy with the unique taxonomic paths in the condensed database. This function returns all matching paths and the corresponding fields from the database. It is important to note that, for the moment, this function is designed to work specifically with taxonomic ranks and assignments from the PR^2 database v 5.0.0 (i.e., for 18S rRNA-based OTUs or ASVs). The result of 'assign_path' can be further improved using the function 'assign_majority', which can annotate taxa with unknown function if enough (e.g., more than 95%) taxa from the parent clade in the focal dataset share a unique functional annotation. Alternatively, the 'assign_sp' function can assign functional information based on species names. Finally, if the taxonomic assignment was performed directly using the GenBank nucleotide database (Sayers et al. 2022) or the SILVA database (Quast et al. 2013), the 'assign_genbank' function accepts a list of GenBank accessions as input.

In addition, the package provides further helper functions, like 'consensing', which allow for the computation of a consensus taxonomy from a set of GenBank accessions (e.g., useful when multiple accessions are associated with a single OTU or ASV) or 'functionize', which condenses any taxonomy data frame based on a specific set of target (functional) fields. It retains only the unique taxonomic paths that share the same target fields. This last function is used internally to create the three condensed versions of the species database, reducing redundancy and facilitating efficient data analysis.

2.5 | Application of EukFunc in a Case Study: The Swiss Western Alps

To showcase the utilisation of the *EukFunc* database and package tools, we re-analysed a dataset of 18S rRNA gene sequences from the Pre-Alps of Western Switzerland published by Seppey et al. (2020, 2023) and Mazel et al. (2021). We demonstrated that

the new possibilities by means of our holistic functional annotation of eukaryotes yield a global understanding of ecosystems by complementing or even providing different functional profiles than those given by protists alone. To test the precision of our approach, we compared the taxonomic ranks at which the protist functional assignments were performed in our study with those in Mazel et al. (2021) based on the same taxonomic annotations. The functional assignments were compared within and between the individual eukaryotic groups (protists, fungi, nematodes), and the correlations with the elevation gradient were investigated.

2.6 | Updating the Database

The current database is built upon the most recent PR^2 v5.0.0 version, a periodically updated database. The updating process requires a manual inspection of the newly added species. The functional information of new species is initially assigned automatically to the clade's unique main and detailed functional information if the taxon belongs to a clade condensed in the *DBc_main* database. On the contrary, the main functional class is set to *unknown* for genera with sister species lacking unique functional information. Subsequently, all new species undergo thorough inspection and validation or correction by taxonomists to ensure accuracy. Once all new entries are characterised, the updated version is published on GitHub. We invite the community of terrestrial eukaryotic biologists to suggest modifications to the database using the issue tracker on the GitHub repository website when an appropriate reference can attest to the validity of the requested change or to provide functional annotation for so far uncovered clades (e.g., mites, springtails, annelids, rotifers). The modification proposal will be assessed by taxonomists regularly, and a new version of the databases will be published (likely once per year), latest, when the previous version has been significantly changed.

3 | Results

3.1 | Database Structure and Functional Groups

The functional reference database comprises up to two main functional categories and contains 14,060 taxa. The two main functional classes for fungi ($N = 8885$) are saprotroph (46.6%) and symbiotroph (51.2%; Figure 1, Table 1). Multiple fungal taxa have a second main functional class (26.6%) that combines either saprotroph and symbiotroph (11.5%) or symbiotroph and saprotroph (15.1%). Only 2.2% of the taxa could not be functionally assigned, falling in the 'unknown' class. The nematodes currently comprise 1402 taxa in the functional database, of which 65.4% are symbiotrophs (mostly parasites) and 34.4% are predators. Multiple nematode taxa are assigned to a secondary functional class, either predator and symbiotroph (0.7%) or symbiotroph and predator (7.5%); only two species (0.1%) cannot be assigned to a main functional class. The functional database currently embraces 5922 protist taxa. Nearly half of these taxa (46.5%) are characterised as phototrophs, the remaining ones are predators (34.6%) or symbiotrophs (16.1%); only 34 taxa are classified as saprotrophs (0.6%). Less than 1.2% of the protists are assigned to a secondary functional class, mainly to the

phototrophs and predators (0.6%) or predators and symbiotrophs (0.4%). Moreover, most fungi have species-specific main functional classes (56.3%), while most nematodes and protist species share their main functional class with congeners (89.2% and 81.8%, respectively; Table 2).

3.2 | Functional Assignment of Western Swiss Alps Soil Eukaryotic Metabarcoding Dataset

The functional annotation of the Swiss Alps soil eukaryotic metabarcoding dataset was achieved by applying the *assign_path() function* on the taxonomic assignment table originally published by Seppey et al. (2020, 2023) with the taxonomic assignment updated to PR^2 V5.0.0 taxonomy, as provided in metaPR2 (Vaulot et al. 2022). Altogether, the dominant function was that of predators (45.2%), followed by that of symbiotrophs (22.6%), saprotrophs (11.3%) and phototrophs (3.6%; Figure 2, Table S5). Fungi were almost evenly distributed between saprotrophs (40.6%) and symbiotrophs (38.3%), while 20.7% lacked a proper affiliation to a functional class. Most nematode taxa were considered predators (72.4%), followed by symbiotrophs (27.6%). Nevertheless, the ratio of predators to symbiotrophs was highly variable in nematodes (Figure S1). Among protists, predators (71.9%) dominated, followed by symbiotrophs (17.2%) and phototrophs (7.1%). The proportions of the protist functional classes were heterogeneous across the samples, with some samples exhibiting a prevalence of phototrophic taxa (Figure S1).

Overall, the functional assignment achieved by means of our database for the protist matched 95% of the original assignments (Table S6). The assignment precision was distinctly improved using the *EukFunc* database with 55.9% and 97.6% of the ASVs assigned at least to order and class levels, respectively, compared to merely 10.1% and 50.9%, respectively, in the original study (Table 3). More than one-third of the ASVs could be assigned functionally at genus rank, and more than 85% of the ASVs assigned to protists, fungi or nematodes could be assigned functionally at class rank or below. Our approach to assigning function to the eukaryome increased by two the number of ASVs analysed per sample, that is, including fungi and nematodes along with protists. For these two newly considered clades, on average, 67% of the ASVs could be functionally assigned per sample.

Elevation-related shifts emerged when the relative abundance of the functional class of all eukaryotes and in each of the three groups (nematodes, protists and fungi) was compared above and below ~2000 m a.s.l. (Figure 3). Different altitudinal patterns among the entire eukaryome and individual eukaryotic functional groups emerged (Figure 3). For example, eukaryotic predators generally increased with elevation, which was driven by protists, while nematodes showed a hump-shaped pattern with strong decreases in predators towards the highest altitudes (Figure 3). In turn, eukaryotic symbiotrophs almost linearly decreased with increased elevation, similar to protists, while fungal symbiotrophs overall remained constant and nematode symbiotrophs showed a U-shaped pattern (Figure 3).

More specifically, within individual eukaryotic groups, protists, phototrophs and predators increased gradually with elevation

TABLE 2 | Number of clades with a unique main functional class or a unique main functional class plus a detailed functional class, and the number of species contained in the respective clades.

| Clade	Level	Main		Detailed	
		Clade	**Species**	**Clade**	**Species**
Fungi	Class	5	19	5	19
	Order	9	36	7	16
	Family	26	130	24	53
	Genus	2255	5748	2172	5045
	Species	2952	2952	3752	3752
Nematodes	Family	195	738	146	332
	Genus	160	621	287	791
	Species	43	43	279	279
Protists	Supergroup	2	13	6	39
	Division	10	517	13	381
	Subdivision	10	511	10	363
	Class	56	1340	61	770
	Order	118	1741	155	2112
	Family	159	696	197	749
	Genus	294	960	726	1612
	Species	144	144	537	537
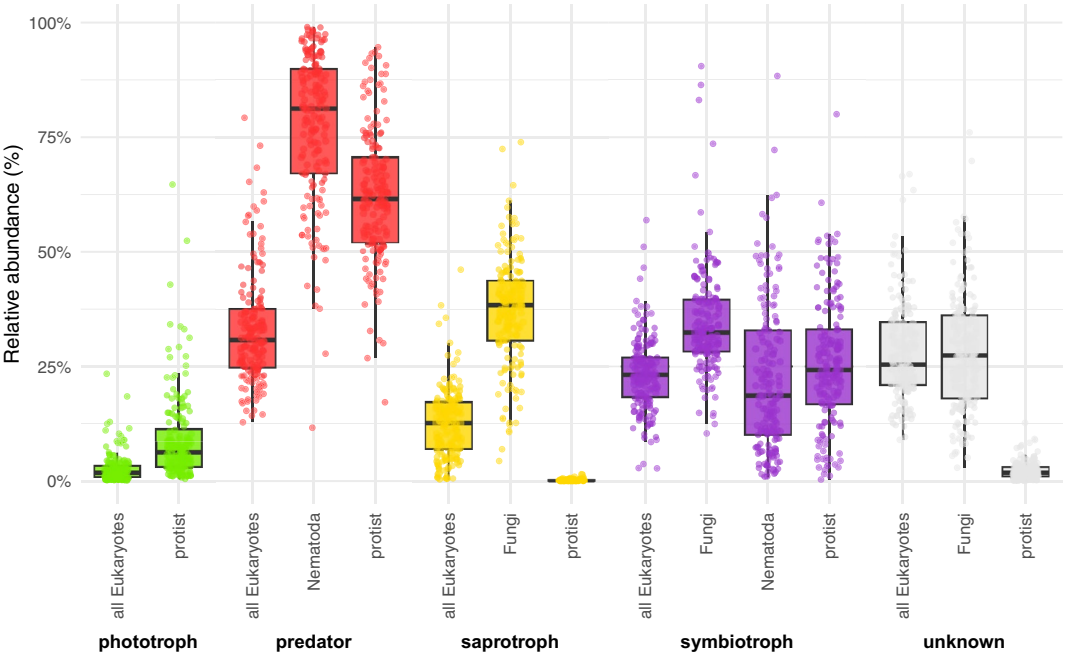
FIGURE 2 | Average relative abundance of each main functional class for all eukaryotic clades considered or each clades (i.e., fungi, nematode and protist) in the Swiss Western Alps dataset.

above 2000 m a.s.l. threshold, whereas the relative proportion of symbiotrophs decreased. For fungi, the saprotrophs peaked around the 2000 m a.s.l. threshold, with some samples composed of more than 60% of saprotrophs. Nematodes displayed a hump-shaped relative abundance distribution along the elevation gradient, with a maximum relative abundance for predators at 2000 m a.s.l.; on the contrary, symbiotrophs displayed higher relative abundances above or below 2000 m a.s.l.

TABLE 3 | Taxonomic rank at which a main functional class was assigned for ASVs of the Swiss Western Alps dataset.

| Clade	Mazel et al. (2021)		This study		This study	
	Protist		Protist		All eukaryotes	
Assigned_at_rank	ASV	Percent ASV	ASV	Percent ASV	ASV	Percent ASV
Supergroup	210	5.2	2	0	2	0
Division	471	11.6	70	1.7	1332	14
Subdivision	1306	32.3	26	0.6	26	0.3
Class	1651	40.8	1687	41.7	1698	17.8
Order	182	4.5	1272	31.4	1341	14
Family	91	2.2	470	11.6	1740	18.2
Genus	129	3.2	458	11.3	2216	23.2
Species	0	0	8	0.2	242	2.5
Unassigned	9	0.2	56	1.4	948	9.9
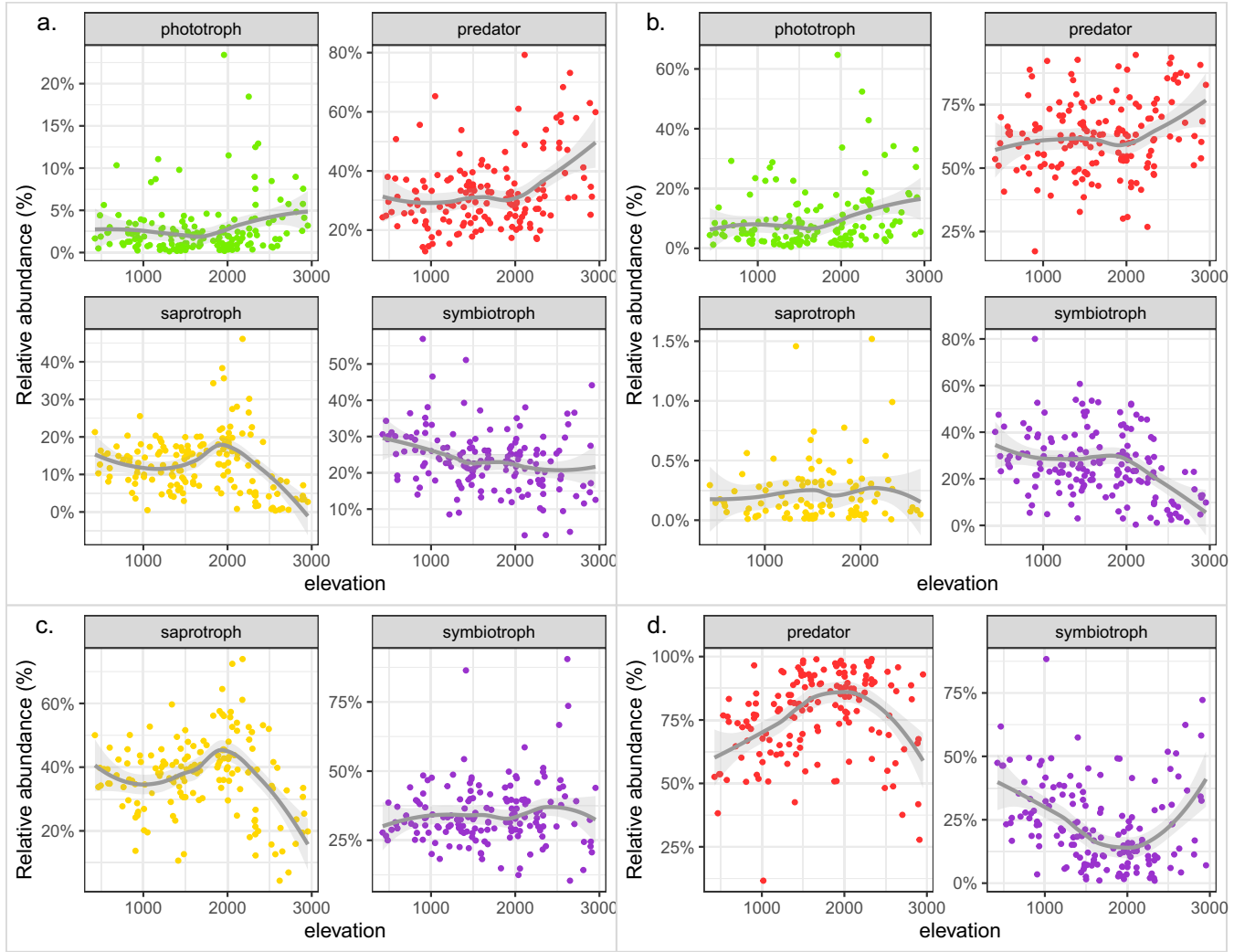
FIGURE 3 | The relative abundances of the main functional classes along the elevation gradient in the Swiss Western Alps dataset. (a) All eukaryotes. (b) Protists. (c) Fungi. (d) Nematodes.

4 | Discussion

To provide a more holistic functional eukaryome database including the numerically dominant soil eukaryotes (fungi, nematodes and protists), we have developed *EukFunc*, providing tools to automatically assign functional classes to OTUs or ASVs obtained in soil metabarcoding studies. This procedure allows a fast annotation of data obtained in soil eukaryote-focused DNA surveys (e.g., based on the 18S rRNA gene) in a reproducible way that allows cross-experiment studies. Most importantly, *EukFunc* allows the combined analysis of the dominant groups of soil eukaryotes, viz., the fungi, nematodes and protists. Together, these taxa represented ca. 95% of all eukaryotic reads in a study encompassing soils from all of Europe (Köninger et al. 2023), and are supposed to constitute a major part of eukaryotic biomass in soil (Bar-On et al. 2018). Our functional eukaryome classification considers most of the microbial primary producers, which is important because soil microbial phototrophs are responsible for 6% of the global net primary production (Jassey et al. 2022). Predation, such as by protists and nematodes, shapes microbiome functioning, leading to changes in the carbon and nutrient cycles, with subsequent impact on plant performance (Topalović and Geisen 2023), while symbiotic mycorrhizal fungi and pathogenic fungi, nematodes and protists also directly determine plant performance (Geisen et al. 2018; Nilsson et al. 2019; Topalović and Geisen 2023). Finally, predation and parasitism might modify other communities by top-down or bottom-up controls (e.g., prokaryotes, aboveground plant communities; Saleem et al. 2013), which influence processes such as the regulation of root secondary metabolite production (Liu et al. 2022).

A future development of our protocol may consider traits that are not related to nutrient acquisition. These traits, like locomotion type, life history and behaviour, might then be used as additional annotation levels for fungi, protists and nematodes. Although these traits are not directly involved in soil nutrient cycling, they can potentially provide valuable information on the ongoing ecological processes in soil (Giachello et al. 2023). However, these traits are usually hard to measure, can be exclusive to particular eukaryotic groups and can fundamentally vary depending on external environmental conditions, thus missing the holistic aim of our project. The accuracy of our pipeline strongly depends on the database size (as for taxonomic assignments; Glöckner et al. 2017), i.e., it will improve gradually as new eukaryotes are added to *EukFunc* (e.g., Gastrotricha, Tardigrada, Rotifera). Special caution must be paid when using the condensed databases, as the aggregated unique main functional assignment of a clade might only be applicable to the most abundant members, while those with different functional classes might be lacking in the database. This is particularly true for protists, for which a significant portion of the main functional classes could be aggregated at the order level and above. Additionally, the V4 or V9 regions of the 18S rRNA gene alone often cannot differentiate organisms at the species level (Lara et al. 2022), hampering the usage of the species-level database; however, for several clades, species identification is achievable (Tragin and Vaulot 2019). This shortcoming of reduced taxonomic depth with shorter sequence information can be overcome by switching to full-length 18S or full-length rDNA operon-based metabarcoding (Gaonkar and Campbell 2024;

Jamy et al. 2020; van Himbeeck et al. 2024). New developments in sequencing technologies now allow for the creation of databases for all eukaryotes and more variable markers, e.g., the ITS region or the full rRNA operon (Tedersoo et al. 2024), which enables higher taxonomic resolution annotations of metabarcoding datasets and thus more precise functional assignments. Researchers can leverage these advancements to enhance the accuracy and resolution of functional assignments in future studies.

As automated sorting of large datasets allows an individual handling of sequences, our protocol outperforms previous studies in the precision with which taxa (based on ASVs or OTUs) can be functionally assigned. Indeed, the immense number of sequences obtained in modern metabarcoding projects forced analysts to subsume functions into big categories built for high taxonomic levels, which resulted in the loss of accuracy in the functional assignment (Glöckner et al. 2017). In protists, for which most 18S rRNA gene-based metabarcoding studies were conducted, previous ones generally failed in the functional annotation of around 10%–15% of the sequences (Mazel et al. 2021; Wang et al. 2021) mainly because of imprecise taxonomic assignment, but also due to missing functional knowledge at higher taxonomic resolution. Similar proportions were inferred in nematodes (Kawanobe et al. 2021). Using single functional assignments at higher taxonomic resolution also causes misassignment, as the main functional class may not be shared even within a genus. Their missing functional assignment below the class level is redressed by our database, which thereby improves both the speed and accuracy of functional annotations.

Using the dataset from (Mazel et al. 2021), we show that the functional profile of the eukaryome is not equally shared by individual members of the eukaryome (fungi, protists and nematodes). Accordingly, studies focusing on individual eukaryotic groups might miss general patterns of the eukaryome (and potentially soil) functioning that are accessible using eukaryote-wide functional placements. In our case study, read numbers of symbiotrophs decreased with elevation, especially above 2000 m a.s.l., matching the findings on parasitic protists (Mazel et al. 2021) (Figure 3). This decrease of protist parasite sequences, combined with the decrease in fungal mycorrhizae and plant-associated nematodes, suggests a general decrease in eukaryotic symbiosis in the alpine regions. In turn, high elevations are characterised by an increased proportion of phototrophs (primary producers) and predators. *EukFunc*, therefore, allows expanding conclusions from single eukaryotic groups to the whole Domain Eukarya, with functional importance at the ecosystem level remaining to be tested experimentally.

To sum up, the application of EukFunc provides a more holistic insight into eukaryome functioning that might deviate from data based on the analyses of individual groups, and likely more reliably informs about soil functions as a wider range of soil biodiversity is covered simultaneously. Our quick and user-friendly method does not require expert knowledge on the soil eukaryome (and individual groups) and is widely applicable to functionally place eukaryotes. Future efforts aim at populating the database with further eukaryotic groups and additional traits to eventually get a more comprehensive understanding of soil systems.

Author Contributions

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The functional database and related R packages functions (R Core Team 2014) are available at https://github.com/lentendu/EukFunc (https://doi.org/10.5281/zenodo.15243078). The code to reproduce the analyses of this article is provided as a Supporting Information.

References

Adl, S. M., D. Bass, C. E. Lane, et al. 2019. "Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes." *Journal of Eukaryotic Microbiology* 66, no. 1: 4–119. https://doi.org/10.1111/jeu.12691.

Anthony, M. A., S. F. Bender, and M. G. A. van der Heijden. 2023. "Enumerating Soil Biodiversity." *Proceedings of the National Academy of Sciences of the United States of America* 120, no. 33: e2304663120. https://doi.org/10.1073/pnas.2304663120.

Baldrian, P., T. Větrovský, C. Lepinay, and P. Kohout. 2022. "High-Throughput Sequencing View on the Magnitude of Global Fungal Diversity." *Fungal Diversity* 114, no. 1: 539–547. https://doi.org/10.1007/s13225-021-00472-y.

Bar-On, Y. M., R. Phillips, and R. Milo. 2018. "The Biomass Distribution on Earth." *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 25: 6506–6511. https://doi.org/10.1073/pnas.1711842115.

Berger, H. 1999. *Monograph of the Oxytrichidae (Ciliophora, Hypotrichia).* Vol. 78. Springer Netherlands. https://doi.org/10.1007/978-94-011-4637-1.

Dumack, K., A. M. Fiore-Donno, D. Bass, and M. Bonkowski. 2020. "Making Sense of Environmental Sequencing Data: Ecologically Important Functional Traits of the Protistan Groups Cercozoa and Endomyxa (Rhizaria)." *Molecular Ecology Resources* 20, no. 2: 398–403. https://doi.org/10.1111/1755-0998.13112.

Faure, E., F. Not, A.-S. Benoiston, K. Labadie, L. Bittner, and S.-D. Ayata. 2019. "Mixotrophic Protists Display Contrasted Biogeographies in the Global Ocean." *ISME Journal* 13, no. 4: 1072–1083. https://doi.org/10.1038/s41396-018-0340-5.

Foissner, W. 1993. *Colpodea (Ciliophora): Vol. 4/1:i–x, 1–798.* Fischer Verlag.

Gaonkar, C. C., and L. Campbell. 2024. "A Full-Length 18S Ribosomal DNA Metabarcoding Approach for Determining Protist Community Diversity Using Nanopore Sequencing." *Ecology and Evolution* 14, no. 4: e11232. https://doi.org/10.1002/ece3.11232.

Geisen, S., E. A. D. Mitchell, S. Adl, et al. 2018. "Soil Protists: A Fertile Frontier in Soil Biology Research." *FEMS Microbiology Reviews* 42, no. 3: 293–323. https://doi.org/10.1093/femsre/fuy006.

Giachello, S., I. Cantera, A. Carteron, et al. 2023. "Toward a Common Set of Functional Traits for Soil Protists." *Soil Biology and Biochemistry* 187: 109207. https://doi.org/10.1016/j.soilbio.2023.109207.

Glöckner, F. O., P. Yilmaz, C. Quast, et al. 2017. "25 Years of Serving the Community With Ribosomal RNA Gene Reference Databases and Tools." *Journal of Biotechnology* 261: 169–176. https://doi.org/10.1016/j.jbiotec.2017.06.1198.

Guillou, L., D. Bachar, S. Audic, et al. 2013. "The Protist Ribosomal Reference Database (PR2): A Catalog of Unicellular Eukaryote Small Sub-Unit rRNA Sequences With Curated Taxonomy." *Nucleic Acids Research* 41, no. Database issue: D597–D604. https://doi.org/10.1093/nar/gks1160.

Jamy, M., R. Foster, P. Barbera, et al. 2020. "Long-Read Metabarcoding of the Eukaryotic rDNA Operon to Phylogenetically and Taxonomically Resolve Environmental Diversity." *Molecular Ecology Resources* 20, no. 2: 429–443. https://doi.org/10.1111/1755-0998.13117.

Jassey, V. E. J., S. Hamard, C. Lepère, et al. 2022. "Photosynthetic Microorganisms Effectively Contribute to Bryophyte CO2 Fixation in Boreal and Tropical Regions." *ISME Communications* 2, no. 1: 64. https://doi.org/10.1038/s43705-022-00149-w.

Kawanobe, M., K. Toyota, and K. Ritz. 2021. "Development and Application of a DNA Metabarcoding Method for Comprehensive Analysis of Soil Nematode Communities." *Applied Soil Ecology* 166: 103974. https://doi.org/10.1016/j.apsoil.2021.103974.

Köninger, J., C. Ballabio, P. Panagos, et al. 2023. "Ecosystem Type Drives Soil Eukaryotic Diversity and Composition in Europe." *Global Change Biology* 29, no. 19: 5706–5719. https://doi.org/10.1111/gcb.16871.

Lara, E., D. Singer, and S. Geisen. 2022. "Discrepancies Between Prokaryotes and Eukaryotes Need to Be Considered in Soil DNA-Based Studies." *Environmental Microbiology* 24, no. 9: 3829–3839. https://doi.org/10.1111/1462-2920.16019.

Lentendu, G., E. Lara, and S. Geisen. 2023. "Metabarcoding Approaches for Soil Eukaryotes, Protists, and Microfauna." In *Microbial Environmental Genomics (MEG)*, edited by F. Martin and S. Uroz, 1–16. Springer US. https://doi.org/10.1007/978-1-0716-2871-3_1.

Liu, Y., D. Li, H. Gao, et al. 2022. "Regulation of Soil Micro-Foodwebs to Root Secondary Metabolites in Cultivated and Wild Licorice Plants." *Science of the Total Environment* 828: 154302. https://doi.org/10.1016/j.scitotenv.2022.154302.

Lynn, D. H. 2008. *The Ciliated Protozoa: Characterization, Classification, and Guide to the Literature.* 3rd ed. Springer.

Mazel, F., L. Malard, H. Niculita-Hirzel, et al. 2021. "Soil Protist Function Varies With Elevation in the Swiss Alps." *Environmental Microbiology* 24, no. 4: 1689–1702. https://doi.org/10.1111/1462-2920.15686.

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. https://doi.org/10.1016/j.funeco.2015.06.006.

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, no. D1: D259–D264. https://doi.org/10.1093/nar/gky1022.

Oliverio, A. M., S. Geisen, M. Delgado-Baquerizo, F. T. Maestre, B. L. Turner, and N. Fierer. 2020. "The Global-Scale Distributions of Soil Protists and Their Contributions to Belowground Systems." *Science Advances* 6, no. 4: eaax8787. https://doi.org/10.1126/sciadv.aax8787.

Patterson, D. J., and J. Larsen, eds. 1992. *The Biology of Free-Living Heterotrophic Flagellates.* Oxford University Press.

Põlme, S., K. Abarenkov, R. Henrik Nilsson, et al. 2020. "FungalTraits: A User-Friendly Traits Database of Fungi and Fungus-Like Stramenopiles." *Fungal Diversity* 105, no. 1: 1–16. https://doi.org/10.1007/s13225-020-00466-2.

Quast, C., E. Pruesse, P. Yilmaz, et al. 2013. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." *Nucleic Acids Research* 41, no. D1: D590–D596. https://doi.org/10.1093/nar/gks1219.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing.

Saleem, M., I. Fetzer, H. Harms, and A. Chatzinotas. 2013. "Diversity of Protists and Bacteria Determines Predation Performance and Stability." *ISME Journal* 7, no. 10: 1912–1921. https://doi.org/10.1038/ismej.2013.95.

Santamaría, J., and G. A. Toranzos. 2003. "Enteric Pathogens and Soil: A Short Review." *International Microbiology* 6, no. 1: 5–9. https://doi.org/10.1007/s10123-003-0096-1.

Sayers, E. W., E. E. Bolton, J. R. Brister, et al. 2022. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 50, no. D1: D20–D26. https://doi.org/10.1093/nar/gkab1112.

Seppey, C. V. W., O. Broennimann, A. Buri, et al. 2020. "Soil Protist Diversity in the Swiss Western Alps Is Better Predicted by Topo-Climatic Than by Edaphic Variables." *Journal of Biogeography* 47, no. 4: 866–878. https://doi.org/10.1111/jbi.13755.

Seppey, C. V. W., E. Lara, O. Broennimann, et al. 2023. "Landscape Structure Is a Key Driver of Soil Protist Diversity in Meadows in the Swiss Alps." *Landscape Ecology* 38, no. 4: 949–965. https://doi.org/10.1007/s10980-022-01572-z.

Sieriebriennikov, B., H. Ferris, and R. G. M. de Goede. 2014. "NINJA: An Automated Calculation System for Nematode-Based Biological Monitoring." *European Journal of Soil Biology* 61: 90–93. https://doi.org/10.1016/j.ejsobi.2014.02.004.

Tedersoo, L., M. S. Hosseyni Moghaddam, V. Mikryukov, et al. 2024. "EUKARYOME: The rRNA Gene Reference Database for Identification of all Eukaryotes." *Database* 2024: baae043. https://doi.org/10.1093/database/baae043.

Topalović, O., and S. Geisen. 2023. "Nematodes as Suppressors and Facilitators of Plant Performance." *New Phytologist* 238, no. 6: 2305–2312. https://doi.org/10.1111/nph.18925.

Tragin, M., and D. Vaulot. 2019. "Novel Diversity Within Marine Mamiellophyceae (Chlorophyta) Unveiled by Metabarcoding." *Scientific Reports* 9, no. 1: 5190. https://doi.org/10.1038/s41598-019-41680-6.

van Himbeeck, R., S. G. Cazzaniga, S. van den Elsen, et al. 2024. "A Full-Length SSU rRNA-Based Workflow for High-Resolution Monitoring of Nematode Communities Reveals Direct and Indirect Responses to Plant-Based Manipulations." *Soil Biology and Biochemistry* 189: 109263. https://doi.org/10.1016/j.soilbio.2023.109263.

Vaulot, D., C. W. H. Sim, D. Ong, et al. 2022. "metaPR2: A Database of Eukaryotic 18S rRNA Metabarcodes With an Emphasis on Protists." *Molecular Ecology Resources* 22, no. 8: 3188–3201. https://doi.org/10.1111/1755-0998.13674.

Wagg, C., S. F. Bender, F. Widmer, and M. G. A. van der Heijden. 2014. "Soil Biodiversity and Soil Community Composition Determine Ecosystem Multifunctionality." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 14: 201320054. https://doi.org/10.1073/pnas.1320054111.

Wang, J., X. Shi, C. Zheng, H. Suter, and Z. Huang. 2021. "Different Responses of Soil Bacterial and Fungal Communities to Nitrogen Deposition in a Subtropical Forest." *Science of the Total Environment* 755: 142449. https://doi.org/10.1016/j.scitotenv.2020.142449.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.