FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild

Nhu H. Nguyen [a, *], Zewei Song [b], Scott T. Bates [b], Sara Branco [c], Leho Tedersoo [d], Jon Menke [a], Jonathan S. Schilling [e], Peter G. Kennedy [a, f]

[a] Department of Plant Biology, University of Minnesota, St. Paul, MN, USA
[b] Department of Plant Pathology, University of Minnesota, St. Paul, MN, USA
[c] Department of Plant & Microbial Biology, University of California Berkeley, Berkeley, CA, USA
[d] Natural History Museum, University of Tartu, Tartu, Estonia
[e] Department of Bioproducts and Biosystems Engineering, University of Minnesota, St. Paul, MN, USA
[f] Department of Ecology, Evolution, & Behavior, University of Minnesota, St. Paul, MN, USA

ARTICLE INFO

ABSTRACT

Fungi typically live in highly diverse communities composed of multiple ecological guilds. Although high-throughput sequencing has greatly increased the ability to quantify the diversity of fungi in environmental samples, researchers currently lack a simple and consistent way to sort large sequence pools into ecologically meaningful categories. We address this issue by introducing FUNGuild, a tool that can be used to taxonomically parse fungal OTUs by ecological guild independent of sequencing platform or analysis pipeline. Using a database and an accompanying bioinformatics script, we demonstrate the application of FUNGuild to three high-throughput sequencing datasets from different habitats: forest soils, grassland soils, and decomposing wood. We found that guilds characteristic of each habitat (i.e., saprotrophic and ectomycorrhizal fungi in forest soils, saprotrophic and arbuscular mycorrhizal fungi in grassland soils, saprotrophic, wood decomposer, and plant pathogenic fungi in decomposing wood) were each well represented. The example datasets demonstrate that while we could quickly and efficiently assign a large portion of the data to guilds, another large portion could not be assigned, reflecting the need to expand and improve the database as well as to gain a better understanding of natural history for many described and undescribed fungal species. As a community resource, FUNGuild is dependent on third-party annotation, so we invite researchers to populate it with new categories and records as well as refine those already in existence.

© 2015 Elsevier Ltd and The British Mycological Society. All rights reserved.

1. Introduction

In recent years, technological advances have allowed researchers to quickly and accurately sequence nearly all microorganisms present in an environmental sample. This has led to a growing appreciation of the highly diverse nature of microbial communities at both local and global scales (Amend et al., 2010; Caporaso et al., 2011; Talbot et al., 2014; Tedersoo et al., 2014). Typically, datasets based on high-throughput sequencing contain millions of sequences and thousands of operational taxonomic units (OTUs). While this level of sequencing depth has increased the ability to successfully capture the richness and composition of microbial communities, researchers are often limited by the inability to effectively parse OTUs into ecologically meaningful categories. In particular, the ability to assign trophic strategies to individual OTUs remains a broad challenge in the field of microbial ecology despite the emerging power of metatranscriptomics (Filiatrault, 2011; Baldrian et al., 2012). Some progress has recently been made for prokaryotes by connecting functional (and often putative) gene families to phylogenetic groups (Langille et al., 2013), but this issue continues to be significant for eukaryotic organisms such as fungi, which, relative to prokaryotes, have large genomes and relatively few fully-sequenced species (Hibbett et al., 2013; Peay, 2014).

The guild concept (also referred to as 'functional group') was created early in ecology (Schimper and Fisher 1902) and broadly

* Corresponding author. 250 Biological Science Building, 1445 Gortner Ave., St. Paul, MN, 55108, USA.
E-mail address: nhnguyen@umn.edu (N.H. Nguyen).

refers to a group of species, whether related or unrelated, that exploit the same class of environmental resources in a similar way (Root, 1967). The fact that guilds have remained in the mainstream of community ecology for decades suggests that it holds a central place in the discipline (Simberloff and Dayan, 1991; Wilson, 1999). Guilds are attractive to ecologists for many reasons, particularly because they provide a way to distill taxonomically complex communities into more manageable ecological units. They also provide a different perspective on community composition than metrics based on species richness or taxonomic identity, due to their focus on trophic strategies. In addition, the use of guilds allows for comparative studies among different communities even when there is no direct overlap in species composition (Hawkins and MacMahon, 1989; Terborgh and Robinson, 1986; Wilson, 1999).

Studies of fungal ecology have, until very recently, focused largely on particular trophic guilds, asking questions only relevant to those guilds while knowingly ignoring the rest of the co-occurring members of the community. For example, it is common for forest ecology studies to focus on just mycorrhizal fungi in a soil sample or decomposers in wood. However, because guilds often interact with each other (Gadgil and Gadgil, 1971; Lindahl et al., 1999; Clemmensen et al., 2015), and in many cases have contrasting responses to the same environmental gradients (Högberg et al., 2003; Kubartová et al., 2012; Štursová et al., 2014), bulk analyses of total fungal communities miss important ecological trends that cancel out or are obscured when considered in aggregate. Some recent high-throughput-based papers have recognized this and begun parsing their sequence pools by guild (Kubartová et al., 2012; Peay et al., 2013; Branco et al., 2013; Štursová et al., 2014; Taylor et al., 2014; Tedersoo et al., 2014; Prober et al., 2015, Clemmensen et al., 2015), recognizing saprotrophs, ectomycorrhizal fungi, plant pathogens, animal pathogens, mycoparasites, wood saprotrophs, lichenized fungi, and even yeasts and dark septate endophytes (the latter two represent morphological groupings rather than trophic guilds). In most cases, however, guild assignment appears to have been made manually.

Guilds can be defined in many ways, with one of the simplest being based on taxonomic affinity (Walter and Ikonen, 1989). In non-fungal systems, this has often been done at the genus level, as demonstrated by the in-depth work of Lambert and Reid (1981), which found desert herptofauna of the same genus frequently belonged to the same guild. There is some evidence that a genus-level cutoff is also applicable for fungi; for instance, the genus *Suillus* is entirely ectomycorrhizal (Tedersoo and Smith, 2013), *Phaenerochaete* are all wood saprotrophs (Burdsall, 1985), and *Puccinia* are all plant pathogens (Cummins and Hiratsuka, 2003). There are, however, notable exceptions, such as the genus *Entoloma*, which contains saprotrophic, ectomycorrhizal and mycoparasitic species, or *Acremonium*, which contains plant pathogen, wood saprotroph and mycoparasitic species. Given the presence of intergeneric variation, species-level identification represents the ideal taxonomic rank at which to assign guilds. Unfortunately, for many fungal genera the ecological lifestyle is known for only a limited number of species. Thus, the application of guild assignments to OTUs at the genus-level (a common phenomenon in the high-throughput studies cited above) would benefit from a measure of confidence to help researchers assess the validity of a given designation.

To address the aforementioned issues, we introduce FUNGuild, a two-component tool consisting of a community-annotated database and a bioinformatics script that parses fungal OTUs into guilds based on their taxonomic assignments. With this tool, we resolve many of the aforementioned challenges as any dataset, containing a few to thousands of OTUs, can be parsed into ecological guilds. Henceforth we will refer to these components as the 'database' and

the 'script' and collectively as "FUNGuild". The name is derived from the concatenation of "Fungi" + "Functional" + "Guild".

2. Methods

2.1. The FUNGuild database and script

FUNGuild v1.0 is a flat database hosted by GitHub (https://github.com/UMNFuN/FUNGuild), making it accessible for use and annotation by any interested party under GNU General Public License. The database currently contains a total of 9476 entries, with 66% at the genus level and 34% at the species level. We have organized entries into three broad groupings refered to as trophic modes (*sensu* Tedersoo et al., 2014): (1) pathotroph = receiving nutrients by harming host cells (including phagotrophs); (2) symbiotroph = receiving nutrients by exchanging resources with host cells; and (3) saprotroph = receiving nutrients by breaking down dead host cells. While these trophic definitions may differ among fields (e.g. pathology vs. ecology), we think that these broadly defined trophic modes work well in fungal community ecology as they reflect the major feeding habits of fungi. Within these trophic modes, we designated a total of 12 categories broadly refered to as guilds (in alphabetical order: animal pathogens, arbuscular mycorrhizal fungi, ectomycorrhizal fungi, ericoid mycorrhizal fungi, foliar endophytes, lichenicolous fungi, lichenized fungi, mycoparasites, plant pathogens, undefined root endophytes, undefined saprotrophs, and wood saprotrophs). The database also includes information on growth morphology, such as yeast, facultative yeast, or thallus (the latter being strictly defined here as a single or aggregation of cells that contributes to a single fungal individual such as some "chytrids" like *Spizellomyces* or members of the Laboulbeniomycetes (Alexopoulos et al., 1996)). Additionally, keywords have been substituted for taxa for cases where taxonomic assignments have returned results such as "uncultured ectomycorrhiza" or "uncultured endophyte", which are only used when taxon assignments are not available. Finally, all entries were assigned a taxonomic level for parsing purposes (see below).

For all database entries a confidence ranking ("highly probable", "probable", and "possible") has been included, reflecting the likelihood that a taxon belongs to a given guild. Whenever possible, confidence assignments were based on assessments given in primary research literature. In rare cases when such information was lacking, other sources were relied on, such as authoritative websites or our own collective research experience (see individual entry annotations in the FUNGuild database for source details). One important issue with both our database and the goal of guild assignment in general is the fact that some fungi do not fall exclusively into a single guild (Promputtha et al., 2007; Parfitt et al., 2010). For example, *Neurospora crassa* may be present as a saprotroph, a pathogen, or an endophyte on plants depending on life stage and environmental conditions (Kuo et al., 2014). In the cases where split ecologies are known, FUNGuild has been set to assign "possible" in the guild confidence category, which effectively weighs each possible ecology equally (see discussion below). This is an important caution about the accuracy of the assignment. While the inclusion of primary literature sources, which accompanies the FUNGuild output, may facilitate more specific interpretations of guild assignment, we emphasize the importance of user awareness about the potential for alternative guild assignments depending on when and where input data was collected.

The script, written in the programming language Python and licensed under GNU General Public License, is compatible with major computer platforms (Windows, Mac, Linux) running Python version 2.7 and above. It works with an input OTU table, formatted

Table 1

An example OTU table after being parsed with FUNGuild. The first four columns (OTU ID, sample 1, sample 2, and taxonomy) are part of the input OTU table and remain unchanged throughout the parsing steps. *The taxonomy here is modified from a UNITE taxonomy string with % identity of sequence attached; it was considerably shortened to fit into the table. Note that although OTU 3 appears to match to the genus *Puccinia*, because the match is below 93%, it should not be considered as a positive match.

OTU ID	Sample 1	Sample 2	Taxonomy*	Trophic mode	Guild	Confidence	Growth morphology	Trait	Notes	Citation/Source
Otu_6	3	8473	99%|JX043001|SH000106.06FU|s__uncultured_ ectomycorrhizal_fungus	Symbiotroph	Ectomycorrhizal	Highly probable				Keyword
Otu_1	0	7456	99%|L54082|SH207619.06FU|s__Suillus_tridentinus	Symbiotroph	Ectomycorrhizal	Highly probable				Rinaldi AC et al., 2008. Fungal Diversity 33: 1—45; Tedersoo L et al., 2010. Mycorrhiza 20: 217—263; http://mycorrhizas.info/ecmf.html
Otu_2	4637	1	95%|EU818897|SH214758.06FU|s__Phanerochaete_sp	Saprotroph	Wood saprotroph	Highly probable		White rot		Glenn JK et al., 1983. Biochemical and Biophysical Research Communications 114: 1077—1083
Otu_5	983	0	100%|FJ596870|SH200896.06FU|s__Entoloma_abortivum	Pathotroph	Mycoparasite	Highly probable			Parasitic on *Armillaria* species	List compiled by John Plischke III; home.comcast.net/~grifola/ fungionfungi.pdf
Otu_4	2	654	96%|UDB008029|SH241227.06FU|s__Entoloma_sp	Symbiotroph| Saprotroph	Ectomycorrhizal| Undefined saprotroph	Possible				Rinaldi AC et al. Fungal Diversity 33: 1—45 (ECM/Not ECM); Tedersoo L et al., 2010. Mycorrhiza 20: 217—263 (sensu stricto, ECM)
Otu_7	0	532	100%|JQ691476|SH229148.06FU|s__Morchella_esculenta	Saprotroph	Undefined saprotroph	Probable				Tedersoo et al. 28 November 2014, Science 346, 1256688
Otu_3	7	30	90%|EF505636|SH220564.06FU|s__Puccinia_sp	Pathotroph	Plant pathogen	Highly probable				Cummins & Hiratsuka. 2003. Illustrated Genera of Rust Fungi, third ed.
Otu_8	0	23	93%|HQ652064|SH274820.06FU|s__Scheffersomyces_insectosa	Saprotroph| Symbiotroph	Wood saprotroph| Animal symbiont	Highly probable	Yeast		Often associated with wood feeding insects	Kurtzman CP et al. (eds.) 2011. The Yeasts, a Taxonomic Study. Fifth Edition. Vols 1—3. Elsevier, San Diego

with OTUs in rows, samples in columns, and a 'taxonomy' column (matching the format of the QIIME pipeline, Caporaso et al., 2010). Under a standard laptop computer, an OTU table with 3000 OTUs can be parsed in under 1 min, and runtime will linearly increase with number of OTUs. The script works by matching terms in the taxonomy column of the OTU table to those in the database, which is called automatically from the GitHub repository. There is potential for each record in the OTU table to have multiple matches due to redundancies in the taxonomic search terms. For example, when searching "*Schizophyllum*", both "*Schizophyllum*" and "*Schizophyllum commune*" would be positive matches as both terms are found as separate items in the database. It is, therefore, necessary for the script to dereplicate the results by keeping the most highly resolved taxon (e.g., species) and discarding all higher-level taxa (e.g., genus, family, and order). As a result, only *Schizophyllum commune* would be kept as the final output. All OTUs not matching taxa in the database are given the designation "unassigned". The output is the original OTU table, sorted by sequence abundance, with "trophic mode", "guild", "confidence", "growth morphology", "trait", "notes", and "reference" columns attached (Table 1). Users can then sort the output file by any guild, split or combine guilds, and decide to keep or discard the data based on additional research after noting the confidence rankings. We suggest that users generally new to studying fungi only accept guild assignments that are "highly probable" or "probable", so as not to overinterpret their data ecologically (however, for the purpose of testing FUNGuild in this study, all of the confidence-ranking assignments were used rather than a stricter use of only "highly probable" and "probable"). A manual for this script along with instructions for database annotation can be found on the GitHub repository link noted above.

2.2. Application of FUNGuild

To demonstrate the application of FUNGuild, three independent high-throughput sequencing datasets of fungi (although FUNGuild works with any OTU table that has a UNITE-based taxonomy, including non-molecular based datasets) were analyzed. Each dataset came from a different research group and originated from different types of environments: forest soils, grassland soils, and decomposing wood. These three datasets were chosen for two primary reasons: (1) to represent a diversity of environments in which researchers may apply the script; and (2) to show that the script is largely independent of upstream issues with sequencing technologies (Shokralla et al., 2012), OTU picking (Lekberg et al., 2014), and taxonomic assignment (Wang et al., 2007; although see discussion below).

The forest study was conducted at the University of Minnesota Cloquet Forestry Center (46°40′45″ N, 92°31′08″ W) in autumn 2013. This site contained six tree species (*Pinus strobus*, *Larix laricina*, *Picea glauca*, *Quercus rubra*, *Betula papyrifera*, and *Acer saccharum*). Four 10 cm deep × 2.5 cm wide cores were taken 1 m apart. Soil samples from each plot were then combined in the field into a single plot sample. The grassland study was conducted at the University of Minnesota Cedar Creek Ecosystem Science Reserve (CCR, 45°24′13″ N, 93°11′20″ W) in autumn 2013. These plots contained a single grass species (*Andropogon gerardii*). Three 10 cm deep × 2.5 cm wide cores were taken from the base of an *A. gerardii* individual. Four replicate plot soil samples were included in the analysis here, and all soil samples were transported in coolers to the laboratory and stored at −20 °C until sieving at 4 mm size prior to DNA extraction. Both datasets were generated using the Illumina MiSeq platform, following Smith and Peay (2014) and Nguyen et al. (2015). The decomposing wood study was also conducted at the University of Minnesota Cloquet Forestry Center (46°42′08″ N, 92°32′53″ W). *Betula papyrifera* logs were cut from healthy trees

and left to decompose on the ground in conifer-dominated sites. Disks were removed from the logs at 7 and 19 months, ground and homogenized, and the fungi within these samples were identified using 454 pyrosequencing. Certain guilds were expected to be well represented in these different environments: undefined saprotrophs in all datasets, ectomycorrhizal fungi in forest soils, arbuscular mycorrhizal fungi in grassland soils, and wood saprotrophs in decomposing logs.

The internal transcribed spacer (ITS) region was used for fungal identification (see Bates et al., 2013; Kõljalg et al., 2013) in all three datasets; however, it is important to note that other gene regions used in sequencing are equally compatible with FUNGuild as it relies on OTU taxonomic assignments rather genetic loci. Sampling methods, bioinformatics quality control, OTU picking and taxonomic assignments were made independently by the respective research groups using slightly different criteria, but all groups produced a final OTU table that is compatible with the FUNGuild script. Guild assignments of OTUs that matched to ≥ 93% sequence similarity to a reference sequence were kept, as this conservatively reflects the genus boundaries with the ITS gene region for many fungi (Nilsson et al., 2008). Since the calling of guilds in FUNGuild is predicated on confidence in the assigned taxonomy (e.g. a sequence identified as *Cenococcum* sp. at 90% sequence similarity may not belong to the genus *Cenococcum*), this 93% threshold is suggested to represent a reasonable general cut-off point for ITS-based inputs. However, genus borders in faster evolving groups may be as low 75% sequence similarity in the ITS region (L. Tedersoo, unpublished data) and phylogenetic analyses should be used to accurately and flexibly delimit genera (see Clemmensen et al., 2013). Users of datasets with more conservative genes, such as the ribosomal large subunit (LSU) and small subunit (SSU), should carefully examine the genus boundaries associated with those genes, as they are not likely to match those based on the ITS region. In addition for clarity, note that this genus-level taxonomy assignment threshold is different than the threshold used in species-level OTU clustering (typically 97%).

3. Results

The forest soil dataset contained 5027113 sequences and 788 OTUs, the grassland soil dataset contained 268577 sequences and 573 OTUs, and the decomposing wood dataset contained 128414 sequences and 134 OTUs. Across these three datasets, FUNGuild detected all three main trophic modes and 10 guilds. Guilds were successfully assigned to 59% of the sequences in the decomposing wood dataset, but only to 32% to both the forest and grassland soils datasets (see the x-axis of Fig. 1). Similarly, guilds were successfully assigned to 58% of the OTUs in the decomposing wood dataset, 40% in the forest soils dataset, and 39% in the grassland soils dataset (see the y-axis of Fig. 1).

Both OTU and sequence richness of guilds varied among datasets. For all three datasets, the unassigned group dominated both OTU richness (41–59%) and sequence richness (40–67%) (Fig. 1A). When unassigned OTUs were removed, undefined saprotrophs were the largest guild in OTU richness (Fig. 1B). The second largest group of guilds included ectomycorrhizal fungi in forest soils, arbuscular mycorrhizal fungi in grassland soils, and wood saprotrophs and plant pathogens in decomposing wood. With regard to sequence richness, the undefined saprotrophs dominated the forest and grassland soils datasets; however, in the decomposing wood dataset, the undefined saprotrophs had only half the number of sequences compared to wood saprotrophs.

Along with the mycorrhizal, saprotrophic, and plant pathogenic guilds, members of four other guilds were also detected in the samples. The undefined root endophytes had higher guild

abundance in the forest and grassland datasets than in the wood dataset. Similarly, ericoid mycorrhizas and animal pathogens were present (albeit in very low abundance) in both the forest and grassland datasets, but absent from the wood dataset. In contrast, mycoparasites and lichenized fungi were both present in the wood dataset, but largely absent from the forest and grassland datasets. Although growth forms such as yeasts are not trophic guilds and thus not included in Fig. 1, they represented a small portion of each dataset with 0.04% OTUs/0.07% sequences in the forest dataset, 0.08% OTUs/0.04% sequences in the grassland dataset, and 0.1% OTUs/0.017% sequences in the wood dataset. All proportions of assigned and unassigned guilds can be found in Table S1.

4. Discussion

Our results showed that FUNGuild was able to efficiently parse three high-throughput sequence pools into ecological meaningful subsets. We chose different datasets with different processing criteria to show that the script is compatible with various sequencing platforms and preprocessing protocols. It is important

to emphasize, however, that guild assignment relies heavily on accurate OTU taxonomic identification. Major databases like Gen-Bank contain a significant number of misidentified sequences (Bidartondo et al., 2008, Nilsson et al., 2006) and "uncultured" sequences, which would place OTUs into the wrong species (and thus potentially the wrong guild) or into the unassigned group. As such, we stress the need for better taxonomy (Gotelli, 2004; Peay, 2014) and integration with well-curated databases (Herr et al., 2014). Because the UNITE database (Abarenkov et al., 2010a; Kõljalg et al., 2013) is continually updated and curated both computationally and by expert input, it currently represents the best database for assigning fungal taxonomy to OTUs generated in high-throughput sequencing studies (Bates et al., 2013). Therefore, we strongly recommend that users of FUNGuild make OTU identifications with the latest version of UNITE reference sequence datasets (https://unite.ut.ee/repository.php) to maximize the chance of correct guild assignment.

In general, we found that the results matched well with our expectations about which guilds would be most represented in each dataset; undefined saptrotrophs and ectomycorrhizal fungi in
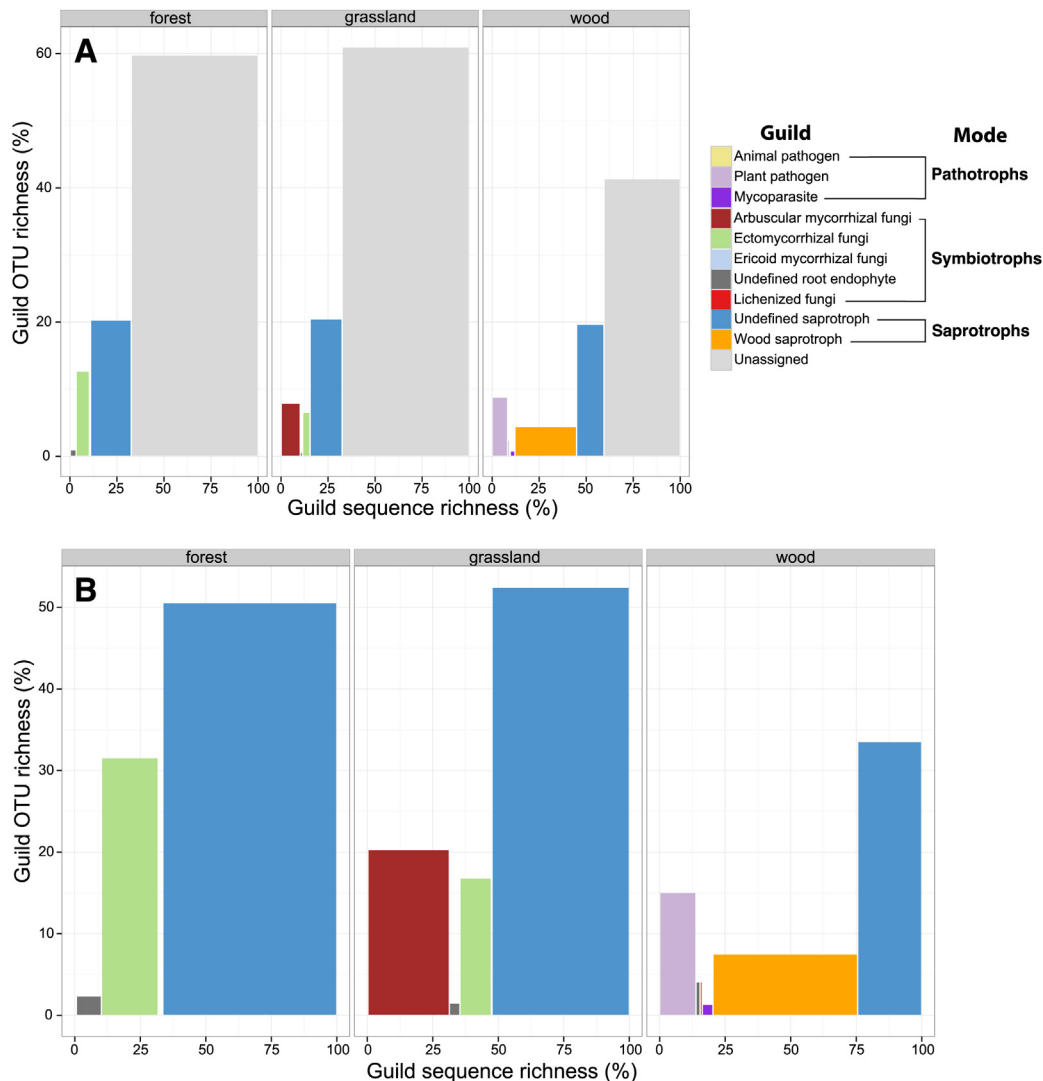

Fig. 1. Guild assignments for three high-throughput environmental datasets using FUNGuild. (A) Proportion of sequence richness (x-axis) and proportion of OTU richness (y-axis) assigned to guilds. OTUs and sequences not assigned to guilds were placed into the unassigned group. (B) subset of the data presented in A, with unassigned OTUs excluded. In the figure, guilds are organized in the same ordering and trophic mode as in the legend.

forest soils (Högberg et al., 2003; Buée et al., 2009; Branco et al., 2013), undefined saprotrophs and arbuscular mycorrhizal fungi in grassland soils (Prober et al., 2015), and undefined saptrotrophs, wood saprotrophs, and plant pathogens in decomposing wood (Kubartová et al., 2012; Ottoson et al., 2014). Although the abundance of OTUs and sequences within each guild can be considered alone, we suggest that combining both dimensions may better reflect the relative importance of a guild in a particular environment (i.e., the area of the bar for each guild in Fig. 1). To highlight this, consider the wood dataset where wood saprotrophs are likely to be the dominant guild. Plant pathogens were twice as OTU-rich as wood saprotrophs, but wood saprotrophs were much more prevalent in terms of sequence abundance. When considered together, the area occupied in Fig. 1 by the wood decomposer guild suggests that it may be more important in terms of ecological effect than plant pathogens for that dataset. We readily acknowledge that many factors affect both sequence abundances (e.g., differences in copy number, primer efficiency, template concentrations, loci sequence length) and OTU abundances (e.g., lineage age, number of taxonomic experts, geographic distribution) (Amend et al., 2010; Nguyen et al., 2015; Tedersoo et al., 2015), so we emphasize that this combined metric would at best provide a starting point to assess potential guild importance. We hope, however, that this approach could be used to inform subsequent research focusing on the impacts different guilds have on their respective environments through the use of more targeted experimental approaches (e.g., Clemmensen et al., 2015).

The ubiquity of the undefined saprotroph guild across datasets was not surprising, given the central role of fungi in decomposition (Dighton, 2003). Finding a dominance of the ectomycorrhizal guild in forest soils was also expected given their known association with forest tree species (Brundrett, 2009), but the considerable presence of this guild in the grassland dataset was not consistent with knowledge about the ecology of those soils. Since these samples were taken > 100 m from any potential tree hosts, it is likely that the ectomycorrhizal sequences found in this habitat came from spores originated in the forests bordering the sampling site (see Koele et al. (2014) for a similar result in arbuscular mycorrhizal forests in New Zealand; Peay et al., 2012). We suspect that the relatively large proportion of plant pathogens in the wood dataset is due to the fact the trees would have likely served as hosts to these plant pathogens before and for sometime after they were felled for the experiment (Sieber, 2007). Additional experimental work to confirm these patterns would help validate the accuracy of FUN-Guild in representing an accurate ecology for the environments studied.

Despite the ability to assign a significant proportion of the OTUs in our three trial datasets to guilds, many OTUs in each dataset were placed into the unassigned group. These findings are similar to Kubartová et al. (2012), who found that the taxonomic richness of "unknowns" (equivalent to our unassigned) made up 41% of their dataset. The 'unassigned' group issue is one that we believe is solvable, but it will require coordinated efforts on multiple fronts. In particular, there appear to be three main barriers to improving the ability of FUNGuild to correctly assign guilds in high-throughput sequence datasets. The first is simply populating the database. Currently, there are only 12 guilds, some of which are well represented (e.g., mycorrhizal fungi and plant pathogens), but others that are not (e.g., foliar endophytes). Thus, the addition of species and genera with known trophic strategies will help to match OTUs that are currently unassigned. In the same way, adding new guild information will also help to decrease the number of unassigned OTUs. For example, in this initial version of FUNGuild, saprotrophs are split only into wood saprotrophs and undefined saprotrophs. Clearly, it would be advantageous to further divide

this major trophic mode into additional guilds. Leaf litter saprotrophs, for example, would be a good target, as those fungi make up an ecologically important group in forests and other ecosystems (Kerekes et al., 2013). We strongly encourage researchers to augment the FUNGuild database with additional annotated records, especially related to information in published studies, and to create new categories where appropriate. Such participation would not only stand to benefit investigators in their work, but it would also be of service to the larger research community who have an interest in fungal ecology. Instructions for annotation of the database are provided through the University of Minnesota Fungal Network GitHub FUNGuild repository (https://github.com/UMNFuN/FUNGuild).

Along with database population, improving the taxonomy of fungal species is also crucial to reduce the large size of the unassigned group and improve the likelihood of guild assignment. Currently, 66% of the entries in the FUNGuild database are at the genus level, which reduces the probability for the best guild assignment as mentioned earlier. Since assigning guilds based on the species level is most accurate, we envision that future versions of the database would ideally contain only species assignments. Given that to date only about 130,000 species of the estimated 6 million species of fungi have been named (www.speciesfungorum.org, Blackwell, 2011; Taylor et al., 2014; but see Tedersoo et al., 2014), improving the efficiency and speed of taxonomic species descriptions will be key to making more complete guild assignments in HTS datasets possible. In the interim, species hypotheses in the UNITE database (Kõljalg et al., 2013) can facilitate assignment of not-yet-described species to guilds. In fact, fungal ecologists can help to integrate the FUNGuild and UNITE databases by including UNITE SH identifiers (i.e. SH208787.07FU) into their published taxonomy metadata. As a recent example, Clemmensen et al. (2015) linked taxonomic information that included UNITE SH identifiers to ecological guild, which made information from that study easy to incorporate into FUNGuild.

A third equally important factor to reducing the number of OTUs that are currently unassigned to guilds is learning more about the autecology of individual fungal species. Doing so will require natural history studies (Tewksbury et al., 2014), where researchers characterize the ecological lifestyles of fungi through integration of environmental metadata, cultivation, and experimentation (see Münzenberger et al. (2012) for an exemplary work). This approach was recently highlighted by Peay (2014), who noted that the power of high-throughput sequencing to move fungal ecology forward is limited without greater effort to characterize species-level natural history. An example comes from the few cultured species of the Archaeorhizomycetes, a recently discovered and described class of fungi commonly found in soil samples at high latitude and altitude sites (Rosling et al., 2011). This group was first identified by environmental sequencing and showed seasonal patterns consistent with plant growth (Schadt et al., 2003), suggesting it may belong to a mycorrhizal guild. However, further studies in pure culture settings and in seedling bioassays indicated that at least the cultured members of the Archaeorhizomycetes are more likely to be root endophytes than mycorrhizal symbionts (Rosling et al., 2011). While studies about fungal life history strategies and trophic modes involve more detailed knowledge about fungi than simple reporting of presence/absence in large environmental sequencing efforts, we stress their critical importance in allowing researchers using high-throughput sequencing approaches to understand the ecology of the communities they are studying.

Like similar tools currently available to fungal and microbial ecologists (e.g. UNITE, SCATA (http://scata.mykopat.slu.se/), PlutoF (Abarenkov et al., 2010b), QIIME (Caporaso et al., 2010)), we have built in the capacity for FUNGuild to grow in terms of coverage and

versatility over time. As mentioned above, the incorporation of additional ecological guilds will continually increase the utility of this tool for the larger research community. For example, additional trait metadata within guilds can also be directly incorporated into FUNGuild. To demonstrate this, we have assigned rot type (white vs. brown, but see Riley et al., 2014) to some taxa within wood decomposer guild to facilitate additional ecological inference in that group (Worrall et al., 1997; Schilling et al., 2015). A similar example involves ectomycorrhizal fungi, which have been increasingly classified according to their extraradical mycelial exploration types (Agerer, 2001; Tedersoo and Smith, 2013). We suggest this additional information would fit well under the "trait" field of the database. Another important future direction involves guild assignments for taxa with multiple guilds. Currently, they are weighted equally, however, future iterations of FUNGuild would benefit by ranking these guilds by their most common occurrence. For example, N. crassa has the confidence of "possible" as a sapro-troph, plant pathogen, and endophyte in this version of FUNGuild, but likely has a "dominant" phase among these three categories. Finally, we also hope that FUNGuild will be readily incorporated in larger bioinformatics workflows commonly used by researchers from different disciplinary backgrounds. The open-source nature of both the database and script should make synchronizing FUNGuild with larger databases and analysis packages (e.g. UNITE, SCATA, PlutoF, QIIME) straightforward. Beyond fungi, the system we have created here could also be broadly applied to guilds in other taxonomic groups, such as nematodes, insects, or basal eukaryotes.

The fungal research community is currently pushing for the unification of sequence-based taxonomy with sequence-based identification and direct connection of sequence-based taxa to their associated metadata (Herr et al., 2014). Here we contribute a step towards this unification by facilitating the connection of taxonomic identification to trophic guilds. FUNGuild provides a way for researchers to comprehensively examine their datasets from an ecological perspective, with an accuracy and speed not feasible without bioinformatics computing power. The example datasets we used demonstrate that while we are able to assign a large portion of the data to guilds, another equally large unassigned portion reflects the need for populating the database, as well as our poor understanding and characterization of the natural history of many described and undescribed fungal species. We, therefore, advocate continued efforts towards the integration of taxonomy, identification, and ecological data, which will collectively push the field of fungal ecology forward.

Acknowledgments

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.funeco.2015.06.006.

References

Abarenkov, K., Henrik Nilsson, R., Larsson, K.-H., et al., 2010a. The UNITE database for molecular identification of fungi—recent updates and future perspectives. New. Phytol. 186, 281—285. http://dx.doi.org/10.1111/j.1469-8137.2009.03160.x.

Abarenkov, K., Tedersoo, L., Nilsson, R.H., et al., 2010b. PlutoF—a web based work-bench for ecological and taxonomic research, with an online implementation for fungal ITS sequences. Evol. Bioinform. 6, 189—196. http://dx.doi.org/10.4137/EBO.S6271.

Agerer, R., 2001. Exploration types of ectomycorrhizae: a proposal to classify ectomycorrhizal mycelial systems according to their patterns of differentiation and putative ecological importance. Mycorrhiza 11, 107—114. http://dx.doi.org/10.1007/s005720100108.

Alexopoulos, C.J., Mims, C.W., Blackwell, M., 1996. Introductory Mycology, fourth ed. Wiley, New York.

Amend, A.S., Seifert, K.A., Bruns, T.D., 2010. Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol. Ecol. 19, 5555—5565. http://dx.doi.org/10.1111/j.1365-294X.2010.04898.x.

Baldrian, P., Kolařík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., Žifčáková, L., Šnajdr, J., Rídl, J., Vlček, Č., Voříšková, J., 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J. 6, 248—258. http://dx.doi.org/10.1038/ismej.2011.95.

Bates, S.T., Clemente, J.C., Flores, G.E., Walters, W.A., Parfrey, L.W., Knight, R., Fierer, N., 2013. Global biogeography of highly diverse protistan communities in soil. ISME J. 7, 652—659. http://dx.doi.org/10.1038/ismej.2012.147.

Bidartondo, M.I., Bruns, T.D., Blackwell, M., Edwards, et al., 2008. Preserving accuracy in GenBank. Science 319, 5870. http://dx.doi.org/10.1126/science.319.5870.1616a.

Blackwell, M., 2011. The Fungi: 1, 2, 3… 5.1 million species? Am. J. Bot. 98, 426—438. http://dx.doi.org/10.3732/ajb.1000298.

Branco, S., Bruns, T.D., Singleton, I., 2013. Fungi at a small scale: spatial zonation of fungal assemblages around single trees. PLoS One 8, e78295. http://dx.doi.org/10.1371/journal.pone.0078295.

Brundrett, M.C., 2009. Mycorrhizal associations and other means of nutrition of vascular plants: understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis. Plant Soil 230, 37—77. http://dx.doi.org/10.1007/s11104-008-9877-9.

Buée, M., Reich, M., Murat, C., Morin, E., Nilsson, R.H., Uroz, S., Martin, F., 2009. 454 pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. New. Phytol. 184, 449—456. http://dx.doi.org/10.1111/j.1469-8137.2009.03003.x.

Burdsall Jr., H.H., 1985. A Contribution to the Taxonomy of the Genus Phanerochaete (Corticiaceae, Aphyllophorales). J. Cramer, Braunschweig.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., et al., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335—336. http://dx.doi.org/10.1038/nmeth0510-335.

Caporaso, J.G., Lauber, C.L., Walters, et al., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. 108, 4516—4522. http://dx.doi.org/10.1073/pnas.1000080107/-/DCSupplemental. www.pnas.org/cgi/doi/10.1073/pnas.1000080107.

Clemmensen, K.E., Bahr, A., Ovaskainen, O., Dahlberg, A., Ekblad, A., Wallander, H., Stenlid, J., Finlay, R.D., Wardle, D.A., Lindahl, B.D., 2013. Roots and associated fungi drive long-term carbon sequestration in boreal forest. Science 339, 1615—1618. http://dx.doi.org/10.1126/science.1231923.

Clemmensen, K.E., Finlay, R.D., Dahlberg, A., Stenlid, J., Lindahl, B.D., 2015. Carbon sequestration is related to mycorrhizal fungal community shifts during long-term succession in boreal forests. New. Phytol. 205, 1525—1536. http://dx.doi.org/10.1111/nph.13208.

Cummins, G.B., Hiratsuka, Y., 2003. Illustrated Genera of Rust Fungi, third ed. APS Press, St. Paul. USA.

Dighton, J., 2003. Fungi in Ecosystem Processes. CRC Press, New York.

Filiatrault, M.J., 2011. Progress in prokaryotic transcriptomics. Curr. Opin. Microbiol. 14, 579—586. http://dx.doi.org/10.1016/j.mib.2011.07.023.

Gadgil, R.L., Gadgil, P.D., 1971. Mycorrhiza and litter decomposition. Nature 233, 133. http://dx.doi.org/10.1038/233133a0.

Gotelli, N.J., 2004. A taxonomic wish-list for community ecology. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 585—597. http://dx.doi.org/10.1098/rstb.2003.1443.

Hawkins, C.P., MacMahon, J.A., 1989. Guilds: the multiple meanings of a concept. Annu. Rev. Entomol. 34, 423—451.

Herr, J.R., Öpik, Maarja, Hibbett, D.S., 2014. Towards the unification of microorganisms. New. Phytol. 205, 27—31.

Hibbett, D.S., Stajich, J.E., Spatafora, J.W., 2013. Toward genome-enabled mycology. Mycologia 105, 1339—1349. http://dx.doi.org/10.3852/13-196.

Högberg, M.N., Bååth, E., Nordgren, A., Arnebrant, K., Högberg, P., 2003. Contrasting effects of nitrogen availability on plant carbon supply to mycorrhizal fungi and saprotrophs — a hypothesis based on field observations in boreal forest. New. Phytol. 160, 225—238. http://dx.doi.org/10.1046/j.1469-8137.2003.00867.x.

Kerekes, J., Kaspari, M., Stevenson, B., Nilsson, R.H., Hartmann, M., Amend, A., Bruns, T.D., 2013. Nutrient enrichment increased species richness of leaf litter fungal assemblages in a tropical forest. Mol. Ecol. 22, 2827—2838. http://dx.doi.org/10.1111/mec.12259.

Koele, N., Dickie, I.A., Blum, J.D., Gleason, J.D., de Graaf, L., 2014. Ecological significance of mineral weathering in ectomycorrhizal and arbuscular mycorrhizal ecosystems from a field-based comparison. Soil Biol. Biochem. 69, 63—70. http://dx.doi.org/10.1016/j.soilbio.2013.10.041.

Kõljalg, U., Nilsson, R.H., Abarenkov, et al., 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22, 5271—5277. http://dx.doi.org/10.1111/mec.12481.

Kubartová, A., Ottosson, E., Dahlberg, A., Stenlid, J., 2012. Patterns of fungal communities among and within decaying logs, revealed by 454 sequencing. Mol. Ecol. 21, 4514—4532. http://dx.doi.org/10.1111/j.1365-294X.2012.05723.x.

Kuo, H.-C., Hui, S., Choi, J., Asiegbu, F.O., Valkonen, J.P.T., Lee, Y., 2014. Secret lifestyles of *Neurospora crassa*. Sci. Rep. 4, 5135. http://dx.doi.org/10.1038/srep05135.

Lambert, S., Reid, W.H., 1981. Biogeography of the Colorado herpetofauna. Am. Midl. Nat. 106, 145—156.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., Mcdonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Thurber, R.L.V., Knight, R., Beiko, R.G., Huttenhower, C., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814—821. http://dx.doi.org/10.1038/nbt.2676.

Lekberg, Y., Gibbons, S.M., Rosendahl, S., 2014. Will different OTU delineation methods change interpretation of arbuscular mycorrhizal fungal community patterns? New. Phytol. 202, 1101—1104.

Lindahl, B., Stenlid, J., Olsson, S., Finlay, R., 1999. Translocation of 32P between interacting mycelia of a wood-decomposing fungus and ectomycorrhizal fungi in microcosm systems. New. Phytol. 144, 183—193. http://dx.doi.org/10.1046/j.1469-8137.1999.00502.x.

Münzenberger, B., Schneider, B., Nilsson, R.H., Bubner, B., Larsson, K.H., Hüttl, R.F., 2012. Morphology, anatomy, and molecular studies of the ectomycorrhiza formed axenically by the fungus *Sistotrema* sp. (Basidiomycota). Mycol. Prog. 11, 817—826. http://dx.doi.org/10.1007/s11557-011-0797-3.

Nguyen, N.H., Smith, D., Peay, K.G., Kennedy, P.G., 2015. Parsing ecological signal from noise in next generation amplicon sequencing. New. Phytol. 205, 1389—1393. http://dx.doi.org/10.1111/nph.12923.

Nilsson, R.H., Kristiansson, E., Ryberg, M., Hallenberg, N., 2008. Intraspecific ITS variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species Identification. Evol. Bioinform. 4, 193—201.

Nilsson, R.H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K.-H., Kõljalg, U., 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS One 1, e59. http://dx.doi.org/10.1371/journal.pone.0000059.

Ottosson, E., Nordén, J., Dahlberg, A., Edman, M., Jönsson, M., Larsson, K.-H., Olsson, J., Penttilä, R., Stenlid, J., Ovaskainen, O., 2014. Species associations during the succession of wood-inhabiting fungal communities. Fungal Ecol. 11, 17—28. http://dx.doi.org/10.1016/j.funeco.2014.03.003.

Parfitt, D., Hunt, J., Dockrell, D., Rogers, H.J., Boddy, L., Griffith, G.W., 2010. Do all trees carry the seeds of their own destruction? PCR reveals numerous wood decay fungi latently present in sapwood of a wide range of angiosperm trees. Fungal Ecol. 3, 338—346. http://dx.doi.org/10.1016/j.funeco.2010.02.001.

Peay, K.G., 2014. Back to the future: natural history and the way forward in modern fungal ecology. Fungal Ecol. 2, 4—9. http://dx.doi.org/10.1016/j.funeco.2014.06.001.

Peay, K.G., Baraloto, C., Fine, P.V.A., 2013. Strong coupling of plant and fungal community structure across western Amazonian rainforests. ISME J. 7, 1852—1861. http://dx.doi.org/10.1038/ismej.2013.66.

Peay, K.G., Schubert, M.G., Nguyen, N.H., Bruns, T.D., 2012. Measuring ectomycorrhizal fungal dispersal: macroecological patterns driven by microscopic propagules. Mol. Ecol. 21, 4122—4136. http://dx.doi.org/10.1111/j.1365-294X.2012.05666.x.

Prober, S.M., Leff, J.W., Bates, S.T., et al., 2015. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. Ecol. Lett. 18, 85—95. http://dx.doi.org/10.1111/ele.12381.

Promputtha, I., Lumyong, S., Dhanasekaran, V., Huge, E., Mckenzie, C., Hyde, K.D., Jeewon, R., 2007. A phylogenetic evaluation of whether endophytes become saprotrophs at host senescence. Microb. Ecol. 53, 579—590. http://dx.doi.org/10.1007/s00248-006-9117-x.

Riley, R., Salamov, A.A., Brown, et al., 2014. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. Proc. Natl. Acad. Sci. 111 http://dx.doi.org/10.1073/pnas.1418116111, 14959—14959.

Root, R.B., 1967. The niche exploitation pattern of the blue-gray gnatcatcher. Ecol. Monogr. 37, 317—350.

Rosling, A., Cox, F., Cruz-Martinez, K., Ihrmark, K., Grelet, G.-A., Lindahl, B.D., Menkis, A., James, T.Y., 2011. Archaeorhizomycetes: unearthing an ancient class of ubiquitous soil fungi. Science 333, 876—879. http://dx.doi.org/10.1126/science.1206958.

Schadt, C.W., Martin, A.P., Lipson, D.A., Schmidt, S.K., 2003. Seasonal dynamics of previously unknown fungal lineages in tundra soils. Science 301, 1359—1361. Http://dx.doi.org/10.1126/science.1086940.

Schilling, J.S., Kaffenberger, J.T., Liew, F.J., Song, Z., 2015. Signature wood modifications reveal decomposer community history. PLOS One 10, e0120679. http://dx.doi.org/10.1371/journal.pone.0120679.

Schimper, A.F.W., Fisher, W.R., 1902. Plant-geography Upon a Physiological Basis. Clarendon Press.

Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. 21, 1794—1805. http://dx.doi.org/10.1111/j.1365-294X.2012.05538.x.

Sieber, T.N., 2007. Endophytic fungi in forest trees: are they mutualists? Fungal Biol. Rev. 21, 75—89. http://dx.doi.org/10.1016/j.fbr.2007.05.004.

Simberloff, D., Dayan, T., 1991. The guild concept and the structure of ecological communities. Annu. Rev. Ecol. Syst. 22, 115—143. http://dx.doi.org/10.1146/annurev.es.22.110191.000555.

Smith, D.P., Peay, K.G., 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. PLoS One 9, e90234. http://dx.doi.org/10.1371/journal.pone.0090234.

Štursová, M., Snajdr, J., Cajthaml, T., Bárta, J., Santrůčková, H., Baldrian, P., 2014. When the forest dies: the response of forest soil fungi to a bark beetle-induced tree dieback. ISME J. 8, 1920—1931. http://dx.doi.org/10.1038/ismej.2014.37.

Talbot, J.M., Bruns, T.D., Taylor, J.W., Smith, D.P., Branco, S., Glassman, S.I., Erlandson, S., Vilgalys, R., Liao, H.-L., Smith, M.E., Peay, K.G., 2014. Endemism and functional convergence across the North American soil mycobiome. Proc. Natl. Acad. Sci. U. S. A. 111, 6341—6346. http://dx.doi.org/10.1073/pnas.1402584111.

Taylor, D.L., Hollingsworth, T.N., McFarland, J.W., Lennon, N.J., Nusbaum, C., Ruess, R.W., 2014. A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. Ecol. Monogr. 84, 3—20. http://dx.doi.org/10.1890/12-1693.1.

Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., Kõljalg, U., Kisand, V., Nilsson, H., Hildebrand, F., Bork, P., Abarenkov, K., 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. MycoKeys 10, 1—43. http://dx.doi.org/10.3897/mycokeys.10.4852.

Tedersoo, L., Bahram, M., Polme, et al., 2014. Global diversity and geography of soil fungi. Science 80. http://dx.doi.org/10.1126/science.1256688.

Tedersoo, L., Smith, M.E., 2013. Lineages of ectomycorrhizal fungi revisited: foraging strategies and novel lineages revealed by sequences from belowground. Fungal Biol. Rev. 27, 83—99. http://dx.doi.org/10.1016/j.fbr.2013.09.001.

Terborgh, J., Robinson, S., 1986. Guilds and their utility in ecology. In: Kikkawa, J., Anderson, D.J. (Eds.), Community Ecology: Pattern and Process. Blackwell, Palo Alto, pp. 65—90.

Tewksbury, J.J., Anderson, J.G.T., Bakker, J.D., et al., 2014. Natural history's place in science and society. Biosci. 64, 300—310. http://dx.doi.org/10.1093/biosci/biu032.

Walter, D.E., Ikonen, E.K., 1989. Species, guilds, and functional groups: taxonomy and behavior in nematophagous arthropods. J. Nematol. 21, 315.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261—5267. http://dx.doi.org/10.1128/AEM.00062-07.

Wilson, J.B., 1999. Guilds, functional types and ecological groups. Oikos 86, 507—522.

Worrall, J.J., Anagnost, S.E., Zabel, R.A., 1997. Comparison of wood decay among diverse lignicolous fungi. Mycologia 89, 199—219.