FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles

Sergei Põlme [1], Kessy Abarenkov [2], R. Henrik Nilsson [3,4], Björn D. Lindahl [5], Karina Clemmensen [6], Havard Kauserud [7], Nhu Nguyen [8], Rasmus Kjoller [9], Scott T. Bates [10], Petr Baldrian [11], Tobias Guldberg Frøslev [12], Kristjan Adojaan [1], Alfredo Vizzini [13], Ave Suija [1], Donald Pfister [14], Hans-Otto Baral [15], Helle Järv [16], Hugo Madrid [17,18], Jenni Nordén [19], Jian-Kui Liu [20], Julia Pawlowska [21], Kadri Põldmaa [1], Kadri Pärtel [1], Kadri Runnel [1], Karen Hansen [22], Karl-Henrik Larsson [23,24], Kevin David Hyde [25], Marcelo Sandoval-Denis [26], Matthew E. Smith [27], Merje Toome-Heller [28], Nalin N. Wijayawardene [29], Nelson Menolli Jr. [30,31], Nicole K. Reynolds [27], Rein Drenkhan [32], Sajeewa S.N. Maharachchikumbura [20], Tatiana B. Gibertoni [33], Thomas Læssøe [34], William Davis [35], Yuri Tokarev [36], Adriana Corrales [37], Adriene Mayra Soares [38], Ahto Agan [1], Alexandre Reis Machado [33], Andres Argüelles- Moyao [39], Andrew Detheridge [40], Angelina de Meiras-Ottoni [33], Annemieke Verbeken [41], Arun Kumar Dutta [42], Bao-Kai Cui [43], C.K. Pradeep [44], César Marín [45,46], Daniel Stanton [47], Daniyal Gohar [1], Dhanushka N. Wanasinghe [48], Eveli Otsing [1], Farzad Aslani [1], Gareth W. Griffith [40], H. Thorsten Lumbsch [49], Hans-Peter Grossart [50,51], Hossein Masigol [52], Ina Timling [53], Inga Hiiesalu [1], Jane Oja [1], John Y. Kupagme [1], József Geml [54], Julieta Alvarez Manjarrez [39], Kai Ilves [1], Kaire Loit [56], Kalev Adamson [32], Kazuhide Nara [55], Kati Küngas [1], Keilor Rojas-Jimenez [57], Krišs Bitenieks [58], Laszlo Irinyi [59,60], Laszlo Nagy [61], Liina Soonvald [32], Li-Wei Zhou [62], Lysett Wagner [40], M.C. Aime [63], Maarja Öpik [1], María Isabel Mujica [64], Martin Metsoja [1], Martin Ryberg [65], Martti Vasar [1], Masao Murata [55], Matthew P. Nelsen [66], Michelle Cleary [67], Milan C. Samarakoon [25], Mingkwan Doilom [68], Mohammad Bahram [1,69], Niloufar Hagh-Doust [1], Olesya Dulya [1], Peter Johnston [70], Petr Kohout [71], Qian Chen [62], Qing Tian [25], Rajasree Nandi [72], Rasekh Amiri [1], Rekhani Hansika Perera [25], Renata dos Santos Chikowski [33], Renato L. Mendes-Alvarenga [33], Roberto Garibay Orijel [39], Robin Gielen [1], Rungtiwa Phookamsak [68], Ruvishika S. Jayawardena [25], Saleh Rahimlou [1], Samantha C. Karunarathna [68], Saowaluck Tibpromma [68], Shawn P. Brown [73], Siim-Kaarel Sepp [1], Sunil Mundra [74,75], Zhu-Hua Luo [77], Tanay Bose [78], Tanel Vahter [1], Tarquin Netherway [79], Teng Yang [80], Tom May [81], Torda Varga [61], Wei Li [82], Victor Rafael Matos Coimbra [33], Virton Rodrigo Targino de Oliveira [33], Vitor Xavier de Lima [33], Vladimir S. Mikryukov [1], Yongzhong Lu [83], Yosuke Matsuda [84], Yumiko Miyamoto [85], Urmas Kõljalg [1,2], Leho Tedersoo [1,2]

1. Institute of Ecology and Earth Sciences, University of Tartu, 14A Ravila, 50411 Tartu, Estonia
2. Natural History Museum, University of Tartu, 14A Ravila, 50411 Tartu, Estonia
3. Gothenburg Global Biodiversity Centre, Box 461, 405 30 Gothenburg, Sweden
4. Department of Biological and Environmental Sciences, University of Gothenburg, Sweden
5. Department of Soil and Environment, Swedish University of Agricultural Sciences, Box 7014, 750 07 Uppsala, Sweden
6. Uppsala BioCenter and Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Box 7026, 750 07 Uppsala, Sweden
7. Section for Genetics and Evolutionary Biology (EvoGene), Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, NO-0316 Oslo, Norway
8. University of Hawai‘i at Mānoa. 3190 Maile Way, St. John 102, Honolulu, Hawai‘i, USA.
9. Department of Biology, University of Copenhagen, Universitetsparken 10, 2100 Copenhagen Ø, Denmark
10. Purdue University Northwest, Westville, IN 46391 USA
11. Institute of Microbiology, Czech Academy of Sciences, Videnska 1083, 14220 Praha 4, Czech Republic
12. GLOBE Institute, University of Copenhagen. Øster Farimagsgade 5, KH7.2.32, DK-1353 Copenhagen, Denmark.
13. Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, Viale Mattioli 25, 10125 Torino, Italy
14. Department of Organismic and Evolutionary Biology and Farlow Library and Herbarium, Harvard University, 22 Divinity Ave., Cambridge, MA 02138 USA
15. Blaihofstr. 42, D-72074 Tübingen, Germany
16. SYNLAB Estonia, Veerenni 53a, Tallinn 11313, Estonia
17. Centro de Genómica y Bioinformática, Universidad Mayor, Camino La Pirámide 5750, Huechuraba, Santiago, Chile
18. Escuela de Tecnología Médica, Universidad Santo Tomás, Los Carreras 753, Osorno, Chile.
19. Norwegian Institute for Nature Research (NINA), Sognsveien 68, 0855 Oslo, Norway
20. School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, 611731, People’s Republic of China
21. Institute of Evolutionary Biology and Biological and Chemical Research Centre, University of Warsaw, ul. Zwirki i Wigury 101, 02-089 Warsaw, Poland
22. Department of Botany, Swedish Museum of Natural History, P.O. Box 50007, SE-10405, Stockholm, Sweden
23. Gothenburg Global Biodiversity Centre, P.O. Box 461, 405 30 Gothenburg, Sweden
24. Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, 0318 Oslo, Norway
25. Center of Excellence in Fungal Research, Mae Fah Luang University, Chiang Rai 57100, Thailand
26. Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands
27. Department of Plant Pathology, University of Florida, FL, USA
28. Plant Health and Environment Laboratory, Ministry for Primary Industries, Auckland 1140, New Zealand
29. Center for Yunnan Plateau Biological Resources Protection and Utilisation, Qujing Normal University, Qujing, Yunnan 655011, China

30. Instituto de Botânica, Núcleo de Pesquisa em Micologia, Av. Miguél Stefano 3687, Água Funda, São Paulo, SP, 04301-012, Brazil
31. Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), Câmpus São Paulo, Rua Pedro Vicente 625, Canindé, São Paulo, SP, 01109-010, Brazil
32. Institute of Forestry and Rural Engineering, Estonian University of Life Sciences, Fr. R. Kreutzwaldi 5, 51006 Tartu, Estonia
33. Departamento de Micologia, Universidade Federal de Pernambuco, Avenida da Engenharia, S/N - Cidade Universitária, 50740-600 Recife, PE, Brazil
34. Department of Biology & Globe Institute, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen Ø, Denmark
35. Oak Ridge Institute for Science and Education, ARS Research Participation Program, Oak Ridge, TN 37830.
36. All-Russian Institute of Plant Protection, Podbelskogo 3 Pushkin, St Petersburg 196608, Russia
37. Department of Biology, Universidad del Rosario, Carrera 24 # 63C-69, Bogota D.C., 111221, Colombia
38. Ciências Biológicas, Universidade Federal Rural da Amazônia, Tomé-Açu, Rodovia PA-451, Km 03, 68.680-000 Bairro Açaizal 68.680-000, Brazil
39. Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico
40. Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Wales, UK
41. Department of Biology, Ghent University, Belgium
42. DST-Inspire Faculty at West Bengal State University
43. Institute of Microbiology, Beijing Forestry University, Beijing 100083, China
44. Jawaharlal Nehru Tropical Botanic Garden & Research Institute (JNTBGRI), Palode, Thiruvananthapuram, Kerala-695562, INDIA
45. Center of Applied Ecology and Sustainability (CAPES), Pontificia Universidad Católica de Chile, 8320000 Santiago, Chile.
46. Institute of Agri-food, Animal and Environmental Sciences (ICA3), Universidad de O'Higgins, 3070000 San Fernando, Chile.
47. Department of Ecology, Evolution and Behavior, University of Minnesota-Twin Cities, Saint Paul 55108 Minnesota, USA
48. Key Laboratory for Plant Biodiversity and Biogeography of East Asia (KLPB), Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China
49. Science and Education and The Grainger Bioinformatics Center, The Field Museum, 1400 S. Lake Shore Dr., Chicago, IL 60605, USA
50. Department of Experimental Limnology, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Alte Fischerhuette 2, D-16775 Stechlin, Germany
51. Institute of Biochemistry and Biology,Potsdam University, Maulbeerallee 2, D-14469 Potsdam, Germany
52. Department of Plant Protection, University of Guilan, Rasht, Iran
53. Institute of Arctic Biology, University of Alaska, Fairbanks, 311 Irving I Building, PO Box 757000, 2140 Koyukuk Drive, Fairbanks, AK, 99775-7000, USA
54. MTA-EKE Lendület Environmental Microbiome Research Group, Eszterházy Károly University, Hungary

56. Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Tartu, Estonia
55. Department of Natural Environmental Studies, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-0882, Japan
57. Escuela de Biologia, Universidad de Costa Rica, 11501 San Jose, Costa Rica
58. Genetic Resource Centre, Latvian State Forest Research Institute "Silava", 111 Rigas str., Salaspils, LV-2169, Latvia
59. Sydney Medical School and Westmead Clinical School and Westmead Institute for Medical Research, University of Sydney, Sydney, NSW, Australia
60. Westmead Hospital, Sydney, NSW, Australia
61. Biological Research Center Szeged, Szeged, Hungary
62. State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China
40. National Reference Center for Invasive Fungal Infections (NRZMyk), Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute (HKI), Jena, Germany
63. Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA
64. Departamento de Ecología, Pontificia Universidad Católica de Chile. Alameda 340, Santiago, Chile.
65. Department of Organismal Biology, Uppsala University, Uppsala, Sweden
66. Science and Education, The Field Museum of Natural History, 1400 S. Lake Shore Dr., Chicago, IL 60605, USA
67. Southern Swedish Forest Research Centre, Swedish University of Agricultural Sciences (SLU), Alnarp, Sweden
68. CAS Key Laboratory for Plant Biodiversity and Biogeography of East Asia (KLPB), Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China
69. Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden
70. Manaaki Whenua - Landcare Research, Private Bag 92170, Auckland 1142, New Zealand
71. Institute of Microbiology of the Czech Academy of Sciences, Videnska 1083, 14220 Praha 4, Czech Republic
72. Institute of Forestry and Environmental Sciences, University of Chittagong, Bangladesh.
73. Department of Biological Sciences, University of Memphis, Memphis, TN, USA
74. Department of Biology, College of Science, United Arab Emirates University, Al-Ain, Abu-Dhabi, UAE
75. Section for Genetics and Evolutionary Biology (EvoGene), Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, NO-0316 Oslo, Norway
77. Key Laboratory of Marine Biogenetic Resources, Third Institute of Oceanography, Ministry of Natural Resources, 184 Daxue Road, Xiamen 361005, China
78. Forestry Agricultural Biotechnology Institute (FABI), Department of Biochemistry, Genetics and Microbiology, University of Pretoria, South Africa.
79. Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden
80. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, East Beijing Road 71, Nanjing 210008, China
81. Royal Botanic Gardens Victoria, Melbourne, Victoria 3004, Australia

82. Ocean University of China, College of Marine Life Sciences, Qingdao,China, 266003
83. School of Food and Pharmaceutical Engineering, Guizhou Institute of Technology, Guiyang 550003, China
84. Graduate School of Bioresources, Mie University, Mie, Japan
85. Arctic Research Center, Hokkaido University, Sapporo, Japan

Declarations

Funding

Financial support was provided by Estonian Science Foundation grants PSG136 and PRG632.

Conflicts of interest/Competing interests

There are no conflicts of interest to declare related to this study.

Availability of data and material

All necessary data is attached and available for public use.

Code availability

Not applicable

Ethics approval

Not applicable

Consent to participate

Not applicable

Consent for publication

All authors give their consent to publish this study in Fungal Diversity.

Abstract

The cryptic lifestyle of most fungi necessitates molecular identification of the guild in environmental studies. Over the past decades, rapid development and affordability of molecular tools have tremendously improved insights of the fungal diversity in all ecosystems and habitats. Yet, in spite of the progress of molecular methods, knowledge about functional properties of the fungal taxa is vague and interpretation of environmental studies in an ecologically meaningful manner remains challenging. In order to facilitate functional assignments and ecological interpretation of environmental studies we introduce a user friendly traits and character database FungalTraits operating at genus and species hypothesis levels. Combining the information from previous efforts such as FUNGuild and Fun^{Fun} together with involvement of expert knowledge, we reannotated 10210 and 151 fungal and Stramenopila genera, respectively. This resulted in a stand-alone spreadsheet dataset covering 17 lifestyle related traits of fungal and Stramenopila genera, designed for rapid functional assignments of environmental studies. In order to assign the trait states to fungal species hypotheses, the scientific community of experts manually categorised and assigned available trait information to 697413 fungal ITS sequences. On the basis of those sequences we were able to summarise trait and host information into 92623 fungal species hypotheses at 1% dissimilarity threshold.

Key words: Fungal traits, Trophic modes, Function, Guild, Bioinformatics, High-throughput sequencing, Community ecology

Introduction

Fungi are one of the most diverse groups of organisms on Earth both in terms of taxonomic richness and functional diversity (McLaughlin & Spatafora 2014; Hawksworth & Lücking 2017). Certain guilds of fungi deliver essential ecosystem functions or colonise habitats too harsh for most other organisms (Mueller et al. 2005; Peay et al. 2016). In particular, lichenised fungi associate with algae or cyanobacteria for energy sources to enable colonisation of inhospitable dry, saline, cold or hot habitats such as nutrient-poor polar and desert soils. Fungal saprotrophs are the most efficient decomposers of dead plant material in soil, water, and sediments (McLaughlin & Spatafora 2014; Grossart et al. 2019). Mycorrhizal fungi promote plant health and nutrition by providing water and nutrients from soil and protection against pathogens, herbivores, and several abiotic stresses (Smith & Read 2008). Besides these unique functions, certain fungi and fungus-like stramenopile groups (Oomycota, Hyphochytriomycota and Labyrinthulida, syn. Labyrinthulomycota) may inhabit plant tissues as endophytes (asymptomatic, commensal or weakly mutualistic inhabitants) and pathogens. Fungal and oomycete pathogens are among the most harmful pests in forestry and agriculture (Hyde et al. 2018; Lucas 2020), whereas oomycetes, labyrinthulids and unicellular fungi of the Aphelidiomycota, Chytridiomycota and Rozellomycota may be the most common parasites of microfauna, protozoans and algae in aquatic habitats (Archibald et al.

2017; Grossart et al. 2019). Fungi are also important agents of disease in animals including humans – especially in immunocompromised patients (Brown et al 2012; de Hoog 2018; Hyde et al. 2018). Although bacteria and viruses are relatively more important parasites of animals (Ryan et al. 2019), chytrids and soil fungi cause the most devastating diseases in amphibians and bats, respectively (Fisher et al. 2012). Because of their capacity to produce antibiotics, toxins and various secondary metabolites, fungi have incredible biotechnological potential, including biocontrol of plant diseases, pests and weeds and stimulation of plant growth (Pavlova et al., 2018; Hyde et al. 2019; Levchenko et al., 2020; Meyer et al. 2020).

Due to the mostly cryptic lifestyles of fungi, molecular methods - especially DNA sequence analysis - have been increasingly used for fungal identification. In the last decade, Sanger-sequencing of amplicons from a single organism has been supplemented by high-throughput sequencing (HTS) methods that enable sequencing of millions of DNA molecules from multiple samples in parallel. This has resulted in unprecedented insights into fungal diversity and taxonomic composition in all types of complex environments, including soil, water, living plant tissues and dust (Nilsson et al. 2018; Větrovský et al. 2020). Curated sequence databases such as UNITE (Nilsson et al. 2019) and ISHAM (Irinyi et al. 2015) have greatly improved our ability to classify fungal operational taxonomic units (OTUs) into species, genera, or higher-level taxa. These OTUs can be compared across samples, studies, and time using the Species Hypothesis (SH) approach, in which species-level proxies are linked to digital object identifiers (DOIs; Kõljalg et al. 2013, 2016).

Thus far, we have limited knowledge about functional properties of most fungi and our insights into their ecology and functions are mainly derived from observational (correlative) field studies. Although most fungi are not easily culturable (Hawksworth 1991), detailed experimental studies are needed to obtain deeper insight into their functions and autecology. Nonetheless, habitat properties provide initial cues about the potential lifestyle of fungal species, especially when isolated from biotrophic structures such as leaf spots or mycorrhizas. However, the trophic mode of many species can be highly variable, including switches between mutualistic, pathogenic and saprotrophic strategies. For example, pathogenic taxa that cause leaf spots may begin as endophytes, but because of environmental stress they become pathogenic and eventually saprotrophs after the death of plant tissues (Promputtha et al. 2005, 2007). The necrotrophic *Rhizoctonia* species (Ceratobasidiaceae) may be pathogens of some plant species, endophytes in others and then also form orchid mycorrhizal symbiosis with Orchidaceae (Veldre et al. 2013). In contrast, the detection of mycorrhizal fungi in clinical samples such as mucosal swabs from patients is suggestive of air-borne propagules or laboratory contamination (Ghannoum et al. 2010). Accordingly, HTS-based taxonomic inventories of fungi provide limited evidence for the functional roles of community members (Nilsson et al. 2018). To overcome these issues, the first panfungal databases linking taxa to lifestyles were published several years ago (Tedersoo et al. 2014; Nguyen et al. 2016), although other, more specific traits databases were already available for, e.g., ectomycorrhizal functional traits (DEEMY; Rambold & Agerer 1997), morphological and chemical traits of lichens (LIAS; Triebel et al. 2007), fruitbody C and N isotope content (Mayor et al. 2009) and fungal sterols (Weete et al. 2010).

Most functional traits of fungi are conserved at the genus-level and sometimes higher taxonomic ranks; therefore, accurate species- or genus-level identification may be used to infer functional traits of taxa (Tedersoo et al. 2014; Nguyen et al. 2016; Zanne et al. 2020). Fungal guilds and taxonomic groups may differ substantially

in patterns of biogeography, community assembly, and host specificity (Tedersoo et al. 2014; Davison et al. 2015; Tisthammer et al. 2016; Põlme et al. 2018). It is also important to separate fungal taxa into guilds or other narrow functional groups for evaluation of associated ecosystem services (Banerjee et al. 2018; Soonvald et al. 2019; Tedersoo et al. 2020). With an increasing number of environmental studies, a large proportion of sequences deposited in repositories are tagged as 'unidentified fungi' or 'uncultured organisms'. These uncertainties, misidentifications and historical synonyms impede proper taxonomic and functional assignments of taxa in HTS studies. Accurate taxonomic assignments in reference data improve taxonomic interpretataion and potentially related functional assignments in environmental studies (Nilsson et al. 2019).

At present, five databases are available for inference of functional guilds and trait information for taxa across most fungal phyla. Tedersoo et al. (2014) published a spreadsheet dataset of fungal guilds and cell types (yeast vs. dimorphic vs. filamentous) based on genus- and family-level information. These data were supplemented and in many places corrected for errors in the FUNGuild database (Nguyen et al. 2016) with assessments of reliability and an important option to allow taxa to be part of multiple guilds simultaneously. FUNGuild also incorporated R and python scripts for automatic assignment of functional guilds to the taxonomic output of HTS bioinformatics platforms. Independently, Jayasiri et al. (2015) presented the FacesOfFungi database that encompasses descriptions of species and genera of mainly Ascomycota. These descriptions include diagnosis as well as ecological, biochemical and economic characterisation, sometimes supplemented with photographs, drawings, and phylogenies. Based on the FacesOfFungi, Guerreiro et al. (2018) developed the ascomycete-specific database 'Notes on genera: Ascomycota' including habitats, substrates, gross biotic interactions and trophic modes. Most recently, Zanne et al. (2020) introduced the Fun^{Fun} database that encompasses much of the data in FUNGuild, supplemented with information about cellular, ecological, and biochemical traits at the species and genus levels. Fun^{Fun} was designed to harbor as many biochemical, genetic, and morphological traits as possible. Similarly to FUNGuild, Fun^{Fun} has a script for assignment of functions to taxa.

Here we present the fungal traits and characters database FungalTraits, a stand-alone spreadsheet dataset, to serve as a resource that provides general ecological information and functional assignment for environmental studies. The framework of FungalTraits was designed during the North European Forest Mycologist (NEFOM) network meeting in Riga, 27-28 November 2014, and elaborated further in subsequent, broader meetings. The main objective was to bridge DNA sequences to the family- and genus-based traits dataset and specimen-related metadata, which have been enabled by recent developments in biodiversity informatics. Supported by a broad international research community, FungalTraits intends to provide comprehensive information about a constrained number of ecologically relevant traits for as many taxa as possible to facilitate trait-based comparative phylogenetics as well as comprehensive analyses in community ecology and macroecology. FungalTraits provides a complementary alternative to existing trait databases, and it seeks to exchange data with other traits databases to provide a rich platform in which to advance fungal biology.

Materials and Methods

Annotations of genera

Starting in September 2012, we systematically compiled ecological information for fungal genera. Parts of this information related to taxa with a sequenced ITS region were published in the reference dataset of Tedersoo et al. (2014) and further in FUNGuild. As of 10 January 2019, we compiled complete lists of all genera of Fungi and fungus-like stramenopiles (the phyla Oomycota, Hyphochytriomycota, and Labyrinthulida) from multiple, largely overlapping sources: Index Fungorum (www.indexfungorum.org), NCBI (www.ncbi.nlm.nih.gov), Mycobank (www.mycobank.org), and the Outline of Fungi beta version (final version in Wijayawardene et al. 2020). We also included numerous synonyms and unused names, because many of these may be revived in forthcoming taxonomic treatments, or applied when sequencing existing collections. Furthermore, there was a substantial conflict among these sources regarding the validity and synonymy of taxon names. Many researchers do not use the accepted names and continue to use taxon names synonymised in some of the data sources but not in others (e.g., *Rhizoctonia* instead of *Ceratobasidium* and *Thanatephorus*, which may have a more specific meaning; Oberwinkler et al. 2013).

In total, we retrieved 10,210 fungal genera and 151 stramenopile genera accepted in at least one of the four sources. For the higher-level taxonomy of Fungi, we followed the Outline of Fungi, which represents a consensus of NCBI, Index Fungorum and Tedersoo et al. (2018a) classifications, and updated this with recently described taxa or new phylogenetic information.

Table 1. Data fields and their properties in genus-level annotation of traits and characters. Numbers in parentheses indicate the number of character states.

Data field	Category	Field type	Importance
primary_lifestyle	selection (30)	guild	primary
secondary_lifestyle	selection (30)/text	guild	secondary
comment_on_lifestyle	text	guild	secondary
plant_pathogenic_capacity	selection (8)	guild	primary
endophytic_interaction_capacity	selection (7)	guild	primary
animal_biotrophic_capacity	selection (19)	guild	primary
decay_substrate	selection (16)	guild	primary

decay_type	selection (8)	guild	secondary
growth_form	selection (15)	body	primary
fruitbody_type	selection (23)	body	primary
hymenium_type	selection (7)	body	secondary
aquatic_habitat	selection (7)	habitat	primary
specific_hosts	text	habitat	secondary
ectomycorrhiza_lineage	selection (87)	specific: ectomycorrhiza	secondary
ectomycorrhiza_exploration_type	selection (7)	specific: ectomycorrhiza	secondary
lichen_primary_photobiont	text	specific: Lichen	secondary
lichen_secondary_photobiont	text	specific: Lichen	secondary

Guild and trait names correspond to the Biological Observation Matrix (BIOM) standards (McDonald et al. 2012), including those proposed in our previous work (Tedersoo et al. 2014; Nguyen et al. 2016; Zanne et al. 2020), with major modifications made to trait arrangement and character (trait) state names (Table 1; Supplementary item 1). To avoid excessive lists of character states and minimise uncertainty in fungal guild assignments, we separated the guild information into separate fields including 'primary_lifestyle', 'secondary_lifestyle', 'plant_pathogenic_capacity', 'endophytic_interaction_capacity', 'animal_biotrophic_capacity', and 'decay_substrate'. The most commonly occurring lifestyle is given in the 'primary_lifestyle' (30 character states). One or more additional lifestyles, if relevant, are given in the 'secondary_lifestyle' field. An additional 'comment_on_lifestyle' field also allows specification of the additional lifestyle, which occurs in only one or a few species, sometimes referring to a particular species. For fully or partly (facultatively) saprotrophic taxa, we generated the fields 'decay_substrate' and 'decay_type', the latter indicating classical decay categories e.g. white rot. However, we anticipate that in nature there is a continuum in the decomposition strategies and large differences within decay types (Riley et al. 2014; Floudas et al. 2020). The 'plant_pathogenic_capacity' field indicates whether plant pathogens occur in this group and which plant groups (e.g. angiosperms, algae, mosses, and liverworts) or organs (leaves, fruits, seeds, roots, etc.)

145 are infected. The ‘endophytic_interaction_capacity’ field indicates whether members of the genus are able to
146 grow as endophytes, classifying these following Rodriguez et al. (2009). The field ‘animal_biotrophic_capacity’
147 enables selection amongst a variety of mutualistic and antagonistic interactions with animals, with further
148 specifications of animal groups (arthropod, coral, fish, invertebrate, vertebrate, termite, and human) and
149 opportunistic interactions with humans.

150 In contrast to other functional databases, we also introduced the trait ‘aquatic_habitat’. This allows categorise
151 fungi as marine, freshwater, more broadly aquatic, or partly water-inhabiting, because many aquatic taxa are
152 often recorded from roots and soil (Chauvet et al. 2016). We find this field of high importance for aquatic
153 studies, as it may be necessary to distinguish accidental spores of terrestrial fungi from taxa that grow naturally
154 inside or on substrates in water (Grossart et al. 2019). We anticipate that such categorisation is subjective for the
155 time being, because we know very little about different life stages of many microfungal genera.

156 We also introduced the traits ‘fruitbody_type’ and ‘hymenium_type’. The ‘fruitbody_type’ covers virtually all
157 classically distinguished types of sexual reproductive structures in Ascomycota, Basidiomycota, and early
158 diverging lineages, and indicates the cases where none are produced. ‘Hymenium_type’ indicates the
159 morphology of hymenium, where the sexual propagules are located relative to the rest of the fruitbody, if
160 relevant. Traditionally, multiple taxonomists and fungal ecologists delimit their research subject by fruitbody
161 type (e.g. polypores, corticioids, agarics, truffles, disco-fungi), although fruitbody and hymenium types are
162 typically not entirely related to other functions. We greatly extended the field ‘growth_form’ that enables 15
163 character states covering amoeboid, filamentous, thalloid, and various unicellular forms relevant to fungi and
164 stramenopiles.

165 We specifically broadened trait information for EcM fungi and lichens. For EcM fungi, we introduced the
166 evolutionary character ‘ectomycorrhiza_lineage’ *sensu* Tedersoo & Smith (2017) and
167 ‘ectomycorrhiza_exploration_type’ following Tedersoo & Smith (2013). Exploration types are defined by the
168 development, form and differentation of extraradical mycelium and rhizomorphs that are related to nutrient
169 acquisition strategies of EcM fungi (Agerer 2001). For exploration types, we used updated information from
170 more recently included or described genera. We furthermore split the short-distance type to short-distance
171 delicate and short-distance coarse, based on the characteristics of emanating hyphae (<1-5 µm diam., thin-
172 walled and cylindrical *vs*. 3-10 µm diam., thick-walled and plump, respectively). Hyphal morphology may
173 indicate the capacity to forage in the immediate vicinity of root tips and preliminary analyses indicate that these
174 two types respond differently to environmental variables and disturbance (Tedersoo et al. 2020). For lichens, we
175 included specific information on the primary and secondary photobiont (as primary_photobiont and
176 secondary_photobiont) obtained from the literature. For non-lichenised taxa, we included the field
177 ‘specific_hosts’ to accommodate information on known exclusive associations with certain taxa (genus- to
178 order-level taxa in Latin, higher taxa in English).

179 As a starting point, we incorporated pre-existing traits information in Tedersoo et al. (2014) and last versions of
180 FUNGuild (accessed 08.10.2018) and FunFun (GitHub dataset; Flores-Moreno et al. 2019; accessed
181 09.12.2019) into FungalTraits. This information was manually parsed into relevant traits fields and reformatted
182 according to our standards to generate a partly filled template. The coauthors with expertise in taxonomy and

fungal ecology were guided to revise the existing information and fill the traits fields with pre-selected character states and add comments to the template whenever relevant. The character states were initially determined by the core group of developers, but several character states were added upon consultation with experts and during data search and insertion. As opinions varied, the terminology represents a compromise among experts. Experts annotated taxa based on their long-term experience with the particular taxonomic groups, scientific literature (e.g., Kurtzman et al. 2011; de Hoog 2019) and specific databases, including LIAS (online version; Rambold et al. 2016), Marine Fungi (Jones et al. 2019) and 'Notes on genera: Ascomycota'. Lists of annotated genera were compared and merged by curators. Nearly half of the genera were covered by two experts. In the rare cases of conflicting annotations, a third opinion was sought. Initially, roughly one-quarter of the genera were covered by no experts (some declined immediately or failed to provide information for various reasons; for several outlying groups no experts were located). Genera belonging to these groups (e.g. Saccharomycotina, Taphrinomycotina, several orders of Sordariomycetes and Leotiomycetes and many small groups) were later annotated by a mycologist with no specific expertise, based on a thorough literature survey, using searches in Google and Google Scholar, the databases FacesOfFungi and Freshwater Ascomycetes Database (Shearer & Raja 2010) as well as comprehensive books (Kurtzman et al. 2011; Pöggeler & Wöstemeyer 2011; McLaughlin & Spatafora 2014; Archibald et al. 2017). The basis of taxonomic and biological knowledge relies on comprehensive work of field taxonomists and plant pathologists. Specific literature sources were not included as citations to speed up the process and avoid dealing with tens of thousands of original references.

Annotation of ITS sequences

To provide trait information to Species Hypotheses, we selected an approach to perform bulk annotation of sequences (including sequenced individuals) from the International Nucleotide Sequence Databases consortium (INSDc) as hosted in the UNITE database (version 7.2). Character states of sequences were merged to individual Species Hypotheses by automated means. Based on the BIOM standards (McDonald et al. 2012), we developed multiple fields for specific traits and character states (Table 2, Supplementary item 1). The data fields included 'DNA_source', with related fields 'culture_source' and 'animal/human_tissue' for specific cases where sequences were obtained directly from pure cultures or animal samples. We included the optional fields 'guild' and 'growth_form' for cases where this was unequivocally clear. In a similar manner, we generated the fields 'ectomycorrhiza_exploration_type', 'ericoid_mycorrhiza_formation', 'endophytic_interaction_capability', 'plant/fungal_pathogenic_capacity', and 'animal/human_biotrophic_interaction_capacity'. Filling these fields presumed that there was sufficient observational or experimental evidence for this indicated in the original commentary field or in the article. The fields 'interacting_taxon' and 'co-occurring_taxon' respectively depict data obtained from specific intimate partners or from a habitat strongly modified by one or more (comma separated) organisms, for example tree species for soil-inhabiting fungi. In both fields, the associated taxon was required to be in Latin, at any taxonomic level. Latin taxon names were checked against the Encyclopedia of Life (www.eol.org) and The Plant List (www.theplantlist.org) for validity and correct spelling. Exclusively for cultures and vouchered specimens, respectively, we added relevant information to the original INSDc data field 'INSD.original.data.Strain' or 'INSD.original.data.Specimen.voucher' that were renamed as 'strain' and 'specimen_voucher' (to meet the BIOM standards implemented in UNITE). We also checked whether these strains or specimens represented type material and added these data to the field 'Type status

(Identification.Typification)' (renamed as 'type_material') when relevant. For geographic information, we combined the existing INSDc data fields into a more standardised format by erecting the fields 'country' (mandatory; to be selected), 'state/province', 'locality_text', 'latitude', 'longitude', 'altitude', 'depth', and 'biome' (to be selected). In addition, we included a general remarks field to specify, e.g., habitat, taxonomy and host for later separation into specific remarks fields related to each main field.

Table 2. Data fields and their properties for sequence-level annotation. Numbers in parentheses indicate the number of character states.

Data Field	Category	Field type	Level
updated_study	study	text (DOI)	mandatory
DNA_source	source	selection (51)	mandatory
culture_source	source	selection (43)	specific
animal/human_tissue	source	selection (30)	specific
guild	guild	selection (23)	individual
growth_form	trait	selection (10)	individual
ectomycorrhiza_exploration_type	trait	selection (6)	individual
ericoid_mycorrhiza_formation	trait	selection (4)	individual
endophytic_interaction_capacity	trait	selection (6)	individual
plant/fungal_pathogenic_capacity	trait	selection (9)	individual
animal/human_biotrophic_interaction_capacity	trait	selection (19)	individual
interacting_taxon	association	text (Latin)	specific

co-occurring_taxa	association	text (Latin)	specific
Strain	collection	text (code)	individual
Specimen_voucher	collection	text (code)	individual
Type_status	collection	selection (16)	individual
Country	locality	selection (243)	mandatory
Sampling_area_State/Province	locality	text	individual
Locality_text	locality	text	individual
Latitude	locality	text (number)	mandatory
Longitude	locality	text (number)	mandatory
Altitude	locality	text (number)	individual
Depth	locality	text (number)	individual
Biome	locality	selection (50)	individual
Remarks	varia	text	individual

In order to assign and summarise trait states of individual records to SHs, we downloaded all ITS sequences by studies and ranked the studies by the number of sequences included, initially focusing only on those with at least 100 sequences. Based on titles, we omitted studies that addressed plants and animals, but included those that covered all eukaryotes. We also excluded studies that produced only ITS1 or ITS2 sequences using HTS techniques, because these subregions separately offer lower taxonomic resolution compared with full-length sequences (Garnica et al. 2016b; Tedersoo et al. 2018b), and are therefore not used for calculating SHs. Nevertheless, the FungalTraits users can still assign short ITS1 and ITS2 reads to SHs (Nilsson et al. 2019). In addition, we searched for potentially high-quality data from environmental studies including <100 sequences,

using the keywords ‘ectomyc’, ‘arbusc’, ‘ericoid’, ‘orchid’ and ‘mycor’ to find studies on mycorrhizae. Similarly, we used the search terms ‘lichen’, ‘endoph’, ‘pathog’, ‘parasit’, ‘root’, ‘aquatic’, ‘water’, and ‘marine’ to capture studies focusing on other guilds or specific underexplored habitat types. We also searched for taxonomic groups focusing on the genera of mycorrhizal fungi and molds as well as stramenopiles. Finally, we searched by names of coauthors to cover their own studies and to allocate annotation tasks to persons directly involved as much as possible. In total, we sought to annotate sequence data in 3,124 studies and unpublished submissions (4.34 % of all submitted datasets) that jointly comprised 414,270 sequences (39.6% of all current fungal ITS sequences in INSDc).

Based on personal contacts and recommendations from other core group members, we invited experts in molecular ecology and phylogenetics to annotate sequence data from 30-50 INSDc studies per expert, with roughly comparable amounts of sequences. The studies were assigned to experts by considering authorship, taxonomic or guild-level expertise and country of origin (in the case of China, India and Iran). The experts received specific instructions for annotations (Supplementary Item 2) and sequence data sorted by studies, including all previous metadata located in multiple fields in the original INSDc format. The experts located and downloaded the studies assigned to them along with supplementary material when relevant. Guided by information in these original studies and INSDc original data fields, experts filled in the data template (Supplementary table 3) following the standards. If the study was marked as unpublished, Google Scholar was used to find the DOI and update relevant study details as far as possible. When mandatory fields could not be filled with information in INSDc or the article, we instructed to contact corresponding authors of these studies. Not surprisingly, contact details of corresponding authors were difficult to find for unpublished studies, especially in the case of authors from China, because of very limited mandatory information for data submission to INSDc. Pointing to personal details, INSDc refused to share contact information of data submitters. In particular, older submissions were hard to track due to digital data decay (Oguz & Koehler 2016).

In addition, FungalTraits curators also annotated data from the remaining studies and submissions by focusing on the fields ‘DNA_source’, ‘interacting_taxon’, ‘co-occurring_taxa’ and ‘country’, using the data previously present in INSDc in non-standard format or misallocated data fields (additional 42,772 submissions comprising 283,173 sequences; 27.1% of INSDc fungal ITS sequences). These data were re-formatted according to our standards. The remaining sequences were mostly short ITS1 or ITS2 reads representing OTUs of HTS sequencing data. All annotated datasets were quality-checked and merged by a curator.

To annotate information about EcM fungal lineages and genera, we downloaded a more recent version 8.2 of UNITE that was released in January 2020. UNITE compound clusters were searched for ectomycorrhizal fungi based on previous information about lineages and named genera. The respective compound clusters were browsed over the PlutoF workbench (Abarenkov et al. 2010) by checking taxonomy and alignments to locate chimeras and low-quality sequences and to update genus-level taxonomy and information about EcM fungal lineages following Tedersoo et al. (2011). Lineage-level and taxonomic assignments were added in a batch mode using the command line.

To assign functional traits for each SH, we included the trait information obtained via annotation of sequences contained within SHs. Because of multiple gaps, complementary and conflicting information in the data, we

used a probabilistic approach that is based on the proportion of specific character states relative to all annotated trait states for each trait per SH. In other words, SHs were considered to possess multiple functional trait states according to the share of these states across sequences.

Results

Genus-level annotations

Across all data fields, the FungalTraits dataset contains 58,479 units (filled cells) of trait information for fungi. We supplemented information about the 'primary_lifestyle' to 8859 out of the 10,210 (86.8%) fungal genera covered (Fig. 1). Other commonly annotated data fields included 'growth_form' (85.9%; Fig. S1) and 'aquatic_habitat' (78.0%; Fig. S1). The 'secondary_guild' and specification to the primary guild ('comment_on_lifestyle') was given for 2280 (22.3%) and 525 (5.1%) genera, respectively. Of the primary lifestyle, wood saprotrophs (19.2%), plant pathogens (15.2%) and litter saprotrophs (11.1%; Fig. 1) were the most common in terms of the number of genera. Saprotrophic, plant pathogenic, endophytic and animal biotrophic capacities occurred in 43.6%, 15.2%, 1.2% and 6.2% of the genera, respectively. Lichenised, ectomycorrhizal, and arbuscular mycorrhizal fungi were assigned to 10.4%, 3.2% and 0.4% of the genera, respectively. Fruitbody and hymenium types were assigned to 70.7% and 69.7% of the genera, respectively (Fig. S1). Of the genera with information on growth form, filamentous (67.2%), thalloid (10.4%), and yeast (4.1%) forms prevailed. Non-aquatic genera clearly dominated (67.7%), followed by freshwater (1.9%) and marine (1.5%) taxa. Altogether 7.4% of the genera included both aquatic and terrestrial species. Specific hosts were assigned to 5.3% of the genera, whereas primary and secondary photobionts were assigned to 9.2% and 0.3% of the genera (88.6% and 2.4% of genera of lichenised fungi as a primary lifestyle), respectively. Nearly all 1209 fungal genera with no trait information were described in the early days of mycology (i.e., before the 1950s) and had no recent information in internet-searchable publications.

With respect to stramenopiles, 682 units of trait information were provided to 150 out of 151 genera. Primary and secondary lifestyles were provided for 93.4% and 20.5% of genera, respectively. Among fungus-like stramenopiles, plant pathogens (29.8%) and animal pathogens (20.5%) prevailed, followed by various saprotrophs (26.5% in total), many of which also have pathogenic potential or include one or more pathogenic species. Information about habitat type was provided for 92.7% of genera; various aquatic habitats combined, terrestrial habitat and partly aquatic habitats taken together characterised 41.7%, 32.5% and 18.5% of genera, respectively (Fig. S2). We added information about growth forms to 94.0% of genera. The distribution of growth forms was strongly related to family- and higher-level taxonomy, with filamentous (rhizomycelial)(22.5%), alternating biflagellate-rhizomycelial (45.7%) and biflagellate (13.2%) forms dominating across all fungus-like stramenopile phyla.

ITS sequence annotations

We assessed metadata for 697,413 INSDc ITS sequences and added data to >85% of these (Table 3). Roughly one-third of the information present in the INSDc dataset was reformatted according to our standards. Information about lineages and genera of EcM fungi were added to >30,000 sequences, whereas 763 sequences were marked as of low-quality or chimeric.

Table 3. The most commonly annotated characters and traits in fungal and stramenopile sequences as based on entry numbers.

	Fungi	Stramenopila
Number of sequences	680,882	16,531
DNA isolation source	565,298	13,791
Country	539,837	13,410
Interacting taxa	302,229	8,628
Biome	145,704	6,328
Guild	77,862	5,158

With respect to isolation source, living culture (20.0%), plant-associated (13.5%), soil (12.9%), and fruitbody (10.6%) were the most common sources of DNA. Furthermore, cultures that were subsequently sequenced, were mostly obtained from plant leaf (17.9%) and other plant-associated material (33.9%; Fig. 2a). Of the interacting taxa, *Homo sapiens* (2.5%), *Fagus sylvatica* (0.8%), and *Glycine max* (0.8%) prevailed (Fig. 2b). The three most commonly annotated interacting guilds were ectomycorrhizal (2.9%), plant pathogens (2.1%), and arbuscular mycorrhizal (1.7%; Fig. S3).

The trait information associated with sequences (Table S2) was integrated into Species Hypotheses (Table S3). The UNITE version 8.2. contains 310,368 eukaryote SHs distinguished at 1% dissimilarity (across 1,799,133 sequences), of which 129,712 SHs (837,572 sequences) represent Fungi and 3,061 SHs (33,566 sequences) are assigned to stramenopiles (including 1,811 SHs and 27,834 sequences of Oomycota). Traits information from the current effort could be assigned to 92,623 (71.4%) of the fungal SHs. Altogether 139,196 (20.0%) out of the total 697,414 sequences for which trait information was added, were not incorporated into any SH because of insufficient length or quality. The most commonly covered SH features included the country of origin

(information available for 95.8% of the SHs), DNA isolation source (92.7%), and interacting taxa (47.2%). Interacting taxa were included as a list of genera and higher-ranking taxa in cases where lower resolution taxonomic data was not available. For example, the top genera *Homo*, *Pinus* and *Quercus* were associated with 4409, 4316, and 3199 SHs, respectively. Conversely, 1546 host genera were associated with a single SH, indicating poor mycological coverage of most plant and animal groups. The largest SH (SH1688425.08; Fig. 3) corresponding to *Alternaria eichhorniae* includes 8,326 sequences, with interacting taxa (68.8% of sequences annotated with such information) belonging to 492 plant genera (including 641 species) and 24 higher-level taxonomic ranks with no genus-level information. This SH was recorded from 88 countries across all continents.

Annotated sequences were assigned to 992 stramenopile SHs at 1% dissimilarity threshold. Of stramenopiles, Oomycota were relatively better annotated (52.2%) than other groups taken together. The three most commonly covered stramenopile characters included DNA isolation source (99.1% of SHs), country of origin (96.0%) and interacting taxa (73.0%). The largest stramenopile SH (SH1791095.08FU) record corresponded to *Phytophthora infestans* that included 748 sequences, associating with five host genera.

Implementation

The genus-level and SH-level annotations represent stand-alone datasets that are available as Table S1 and Table S3. The current version and future versions of FungalTraits can be downloaded from the UNITE homepage (https://unite.ut.ee/repository.php). We intend to release a new FungalTraits version when UNITE SHs are updated. The original annotations of genera and sequences remain attached to the genus names and sequence accession numbers, respectively. In new versions, the proportions of trait states will be re-calculated. For genus names, we do not consider any automatic procedure when these are synonymised or split into new genera. We intend to consider annotations to newly described genera when major changes in taxonomy occur or within at least five years.

Assignment of guild and trait information from FungalTraits to custom ecological or phylogenetic datasets can be accomplished in several ways. Both genus-level and SH-level traits are available in a ready-to-use comma-separated value (.csv) text format. The *vlookup* function in MS Excel and similar functions in other spreadsheet programs enable rapid assignment of functional trait states to genus names and SH codes in the taxonomic identification tables and OTU tables produced as an output in nearly all HTS bioinformatics workflows. An example of using the *vlookup* function is given in Table S4.

To test the performance of FungalTraits, we used a global dataset of 50,589 OTUs (21,468 OTUs determined at the genus level; Tedersoo et al. 2014). Setting up the *vlookup* function for all fields and obtaining results took 9 minutes on a desktop PC. The same dataset took a roughly comparable amount of processing time for FUNGuild using the python script (Table S5).

Discussion

One of the main criticisms of HTS-based metabarcoding studies is that only diversity is assessed without addressing functional differences among taxa (Hongsanan et al. 2018; Nilsson et al. 2018; Zanne et al. 2020). Carefully curated sequence and taxon references would substantially benefit ecological interpretations of HTS studies (Nilsson et al. 2018; Lücking et al. 2020). We have therefore developed a new combined approach for genus- and SH-level trait annotation to promote functional information assignment to fungi and fungus-like stramenopiles in ecological and evolutionary research. Based on literature and taxonomic expertise, nearly all actively used fungal genera were functionally annotated to some extent, which doubles previous efforts in taxonomic breadth and increases the number of data points by an order of magnitude. Similarly, the standardised metadata added to sequences exceeds our previous effort a decade ago (Tedersoo et al. 2011) by an order of magnitude. To our knowledge, the process of calculating proportional traits based on individual sequences and sequenced individuals in species-level taxa (SHs) is entirely novel. This information complements the genus-level annotations of taxa with contrasting lifestyles or interacting taxa. Furthermore, SH-based trait annotation greatly adds to information about geographic distribution and ecology for fungal taxa that cannot be reliably assigned to any known genus or higher ranking taxon.

Although the ecological traits of fungi are typically conserved at the genus level and sometimes at higher taxonomic levels (Zanne et al. 2020), there are multiple occasions where the same species has diverse functions or members of the same genus display different trophic strategies and fall into different functional guilds (Nguyen et al. 2016; Selosse et al. 2018). Besides the primary guild, which is expected to be the most characteristic to particular genera, we, therefore, generated extra fields for these secondary functions (including specification for particular species) and capacities to perform certain biotrophic functions such as the ability to perform as plant pathogens, endophytes, saprotrophs, or animal biotrophs including opportunistic parasitism in animals and humans. These fields enable researchers focused on specific objectives to find answers relevant to their questions more efficiently. Considering the needs of fungal and plant ecologists, we also added information about specific interacting taxa, reproductive structures, fruitbody form, and the capacity to inhabit aquatic environments. A majority of these trait fields are not covered in other fungal traits databases such as FacesOfFungi, FUNGuild and Fun^{Fun}. Because our objective was to focus on a relatively small number of ecological traits with comprehensive taxonomic coverage, other databases may be better suitable for finding alternative or more specific information. For example, we recommend researchers to visit the FacesOfFungi and Marine Fungi databases for more species-level information about morphological characters and habitat. FUNGuild has the associated assignment probability field and comprehensive remarks about taxa with multiple lifestyles. FunFun gives an overview of most fungal traits recorded so far and it provides complementary information about a number of morphological (spores), biochemical (enzymes), geographic (known distribution) and genome-encoded (presence of certain metabolism-related genes) traits not covered by the first version of FungalTraits. FunFun database is designed to work in the R environment and it can be also used as a stand-alone database to perform quick searches.

Because of the simple .csv spreadsheet format, FungalTraits can be used without skills in R or python environments and it requires no dataset formatting prior to analyses. By selecting relevant data fields for the output, FungalTraits enables a custom choice of trait fields and it requires no skills in the use of R or python environments. The spreadsheet format is also helpful for rapid manual searches to check available information

for selected taxa for any educational purpose. The functional assignment algorithms of all three databases are fully reproducible.

All taxonomic and trait assignment tools require final decision-making by users, considering the intended taxonomic resolution, relevant functional groups, and sources. First, users should consider a suitable clustering approach and sequence similarity threshold for distinguishing OTUs or SHs and genera (Lindahl et al. 2013; Nilsson et al. 2018). Depending on molecular markers and taxonomic groups, critical sequence dissimilarity thresholds for species and genus levels may vary (e.g. Garnica et al. 2016). The same applies to a selection of a proper SH level (Kõljalg et al. 2013). It is also important to bear in mind that taxonomic and functional assignments should be conducted at appropriate taxonomic levels - as a rule of thumb, the sequence differences of the obtained OTUs to reference SHs should not exceed the clustering threshold.

For functional annotation of organisms, species-level assignments are certainly the most precise, but there are several technical obstacles for generating species-level functional databases. First, there is a huge number of described species, the amount of which can only be handled by thousands of experts (Hawksworth & Lücking 2017). Second, DNA sequence data suggest that a large proportion of morphological species are actually represented by several or even hundreds of molecular species that may conform better to the biological species criterion (Taylor et al. 2006; Lücking et al. 2014). Therefore, multiple SHs commonly represent the same morphological species. Typically, valid taxon names cannot be ascribed to a single SH, because the type specimen is not sequenced or information about this is missing. Alongside with previous efforts (Nilsson et al. 2014: Schoch et al. 2014), we annotated type status to sequences representing type material, to be able to integrate traits information and other metadata with valid species names. In the future versions of FungalTraits, we intend to merge the taxonomy-based and sequence-based approaches by operating more on a species/SH level and focus on species that have distinct traits compared with those characteristic of the rest of the genus.

To conclude, fungal traits data are increasingly used by ecologists, as judged from the number of citations to pioneer studies. Therefore, we propose FungalTraits - a global research community-supported, easy-to-use functional traits database that covers multiple newly compiled traits and a large proportion of fungal and stramenopile taxa as well as their published sequences incorporated in Species Hypotheses. The straightforward spreadsheet format of the data provides easy data exchange options with other databases. In the future, we intend to establish the connection between SHs and species, so that it is possible to integrate traits derived from molecular identification and metadata with those derived from microscopic and –omics studies of specimens. Experts and users who wish to update or revise species- and genus-level traits and character states are guided to the online spreadsheet document at URL. These suggestions are revised by curators and implemented in the next version of FungalTraits.

Acknowledgments

Financial support was provided by Estonian Science Foundation grants PSG136 and PRG632. We thank Dr. Mario Saare for help with the R software.

References

Abarenkov K, Tedersoo L, Nilsson RH, Vellak K, Saar I, Veldre V, Parmasto E, Prous M, Aan A, Ots M, Kurina O, Ostonen I, Jõgeva J, Halapuu S, Põldmaa K, Toots M, Truu J, Larsson K-H, Kõljalg U (2010) PlutoF – a web based workbench for ecological and taxonomic research with an online implementation for fungal ITS sequences. Evol Bioinform 6:189-196

Agerer R (2001) Exploration types of ectomycorrhizae. Mycorrhiza 11:107-114

Archibald JM, Archibald GB, Slamovits CH (eds.) (2017) Handbook of the Protists. Springer, Cham.

Banerjee S, Schlaeppi K, van der Heijden MG (2018) Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol 16:567

Brown GD, Denning DW, GowNA, Levitz SM, Netea MG, White TC (2012) Hidden killers: human fungal infections. Sci Transl Med 165:165rv113

Chauvet E, Cornut J, Sridhar KR, Selosse M-A, Bärlöcher F (2016) Beyond the water column: aquatic hyphomycetes outside their preferred habitat. Fung Ecol 19:112-127

Davison J, Moora M, Öpik M, Adholeya A, Ainsaar L, Ba A, Burla S, Diedhiou AG, Hiiesalu I, Jairus T, Johnson NC, Kane A, Koorem K, Kochar M, Ndiaye C, Pärtel M, Reier Ü, Saks Ü, Singh R, Vasar M, Zobel M (2015) Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. Science 28:970-973

De Hoog GS, Guarro J, Gené J, Figueras MJ (2018) Atlas of clinical fungi, 4th edn. Westerdijk Institute/Universitat Rovira i Virgili, Utrecht/Reus

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ (2012) Emerging fungal threats to animal, plant and ecosystem health. Nature 484:186–194

Flores-Moreno H, Treseder KK, Cornwell WK, Maynard DS, Milo A, Abarenkov K, Afkhami ME, Aguilar-Trigueros CA, Bates S, Bhatnagar JM, Busby PE, Christian N, Crowther TW, Floudas D, Gazis R, Hibbett D, Kennedy PF, Lindner DL, Nilsson RH, Powell J, Schildhauer M, Schilling J, Zanne AE (2019) fungaltraits aka funfun: a dynamic functional trait database for the world's fungi. https://github.com/traitecoevo/fungaltraits

Floudas D, Bentzer J, Ahrén D, Johansson T, Persson P, Tunlid A (2020) Uncovering the hidden diversity of litter-decomposition mechanisms in mushroom-forming fungi. ISME J 14:2046–2059

Garnica S, Schön ME, Abarenkov K, Riess K, Liimatainen K, Niskanen T, Dima B, Soop K, Frøslev TG, Jeppesen TS, Peintner U (2016) Determining threshold values for barcoding fungi: lessons from *Cortinarius* (Basidiomycota), a highly diverse and widespread ectomycorrhizal genus. FEMS Microbiol Ecol 92:fiw045

Ghannoum M, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM (2010) Characterisation of the oral fungal microbiome (Mycobiome) in healthy individuals. PLoS Path 6:e1000713

Grossart HP, Van den Wyngaert S, Kagami M, Wurzbacher C, Cunliffe M, Rojas-Jimenez K (2019) Fungi in aquatic ecosystems. Nat Rev Microbiol 17:339–354

Guerreiro MA, Wijayawardene NN, Hyde KD, Peršoh D (2018) – Ecology of Ascomycete genera – A searchable table of "Notes on genera: Ascomycota". Asian J Mycol 1:146–150

Hawksworth DL (1991) The fungal dimension of biodiversity: magnitude, significance, and conservation. Mycol Res 95:641-655

Hawksworth DL, Lücking R (2017) Fung Divers revisited: 2.2 to 3.8 million species. Microbiol Spectr 1:79-95.

Hongsanan S, Jeewon R, Purahong W, Xie N et al. (2018) Can we use environmental DNA as holotypes? Fung Divers 92:1-30

Hyde KD, Al-Hatmi AM, Andersen B, Boekhout T, Buzina W, Dawson TL, Eastwood DC, Jones EG, de Hoog S, Kang Y, Longcore JE (2018) The world's ten most feared fungi. Fung Divers 93:161–194

Hyde KD, Xu JC, Lumyong S, Rapior S et al. (2019) The amazing potential of fungi, 50 ways we can exploit fungi industrially. Fung Divers 97:1–136

Irinyi L, Serena C, Garcia-Hermoso D, Arabatzis M, Desnos-Ollivier M, Vu D (2015) International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database—the quality controlled standard tool for routine identification of human and animal pathogenic fungi. Med Mycol 55:313-337

Jayasiri SC, Hyde KD, Ariyawansa HA, Bhat J, Buyck B, Cai L, Dai YC, Abd-Elsalam KA, Ertz D, Hidayat I, Jeewon R, Jones EBG, Bahkali AH, Karunarathna SC, Liu JK, Luangsa-ard JJ, Lumbsch HT, Maharachchikumbura SSN, McKenzie EHC, Moncalvo, JM, Ghobad-Nejhad M, Nilsson H, Pang KA, Pereira OL, Phillips AJL, Raspé O, Rollins AW, Romero AI, Etayo J, Selçuk F, Stephenson SL, Suetrong S, Taylor JE, Tsui CKM, Vizzini A, Abdel-Wahab MA, Wen TC, Boonmee S, Dai DQ, Daranagama DA, Dissanayake AJ, Ekanayaka AH, Fryar SC, Hongsanan S, Jayawardena RS, Li WJ, Perera RH, Phookamsak R, de Silva NI, Thambugala KM, Tian Q, Wijayawardene NN, Zhao RL, Zhao Q, Kang JC, Promputtha I (2015) The Faces of Fungi database: fungal names linked with morphology, phylogeny and human impacts. Fung Divers 74:3-18

Jones EG, Pang KL, Abdel-Wahab MA, Scholz B, Hyde KD, Boekhout T, Ebel R, Rateb ME, Henderson L, Sakayaroj J, Suetrong S (2019) An online resource for marine fungi. Fung Divers 96:347-433

Kurtzman CP, Fell JW, Boekhout T (eds.). 2011. The Yeasts, a Taxonomic Study. Elsevier, Amsterdam.

Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M (2013) Towards a unified paradigm for sequence-based identification of Fungi. Mol Ecol 22:5271–5277

Kõljalg U, Tedersoo L, Nilsson RH, Abarenkov K (2016) Digital identifiers for fungal species. Science 352:1182-1183

Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjøller R, Kõljalg U, Pennanen T, Rosendahl S, Stenlid J, Kauserud H (2013) Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. New Phytologist 199:288–299

Levchenko MV, Kononchuk AG, Gerus AV, Lednev GR (2020) Differential susceptibility of *Locusta migratoria* and *Schistocerca gregaria* (Orthoptera: Acrididae) to infection with entomopathogenic fungi. Plant Protect News 103:150-152

Lücking R, Aime MC, Robbertse B, Miller AN, Ariyawansa HA, Aoki T, Hawksworth DL (2020) Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal DNA barcoding? IMA Fung 111:1-32

Lücking R, Dal-Forno M, Sikaroodi M, Gillevet PM, Bungartz F, Moncada B, Yánez-Ayabaca A, Chaves JL, Coca LF, Lawrey JD (2014) A single macrolichen constitutes hundreds of unrecognised species. Proc Natl Acad Sci USA 111:11091-11096

516 Mayor JR, Schuur EAG, Henkel TW (2009) Elucidating the nutritional dynamics of fungi using stable isotopes.
517 Ecol Lett 12:171-183
518 McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Knight R (2012) The Biological
519 Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. GigaScience
520 1:7
521 Meyer V, Basenko EY, Benz JP, Braus GH, Caddick MX, Csukai M, De Vries RP, Endy D, Frisvad JC, Gunde-
522 Cimerman N (2020) Growing a circular economy with fungal biotechnology: A white paper. Fung Biol
523 Biotechnol 7:1–23
524 Mueller UG, Gerardo NM, Aanen DK, Six DL, Schultz TR (2005) The evolution of agriculture in insects. Annu
525 Rev Ecol Evol Syst 36:563-595
526 Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG (2016) FUNGuild:
527 an open annotation tool for parsing fungal community datasets by ecological guild. Fung Ecol 20:241–248
528 Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L (2018) Mycobiome diversity: high-
529 throughput sequencing and identification of fungi. Nat Rev Microbiol 17:95–109
530 Nilsson RH, Hyde KD, Pawłowska J, Ryberg M, Tedersoo L, Aas AB, Alias SA, Alves A, Anderson CL,
531 Antonelli A, Arnold AE, Bahnmann B, Bahram M, Bengtsson-Palme J, Berlin A, Branco S, Chomnunti P,
532 Dissanayake A, Drenkhan R, Friberg H, Frøslev TG, Halwachs B, Hartmann M, Henricot B, Jayawardena R,
533 Jumpponen A, Kauserud H, Koskela S, Kulik T, Liimatainen K, Lindahl BD, Lindner D, Liu J-K,
534 Maharachchikumbura S, Manamgoda D, Martinsson S, Neves MA, Niskanen T, Nylinder S, Pereira OL,
535 Pinho DB, Porter TM, Queloz V, Riit T, Sánchez-García M, Sousa FD, Stefańczyk E, Tadych M, Takamatsu
536 S, Tian Q, Udayanga D, Unterseher M, Wang Z, Wikee S, Yan J, Larsson E, Larsson K-H, Kõljalg U,
537 Abarenkov K (2014) Improving ITS sequence data for identification of plant pathogenic fungi. Fung Divers
538 67:11-19
539 Nilsson RH, Larsson KH, Taylor AF, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K,
540 Glöckner FO, Tedersoo L, Saar I. Kõljalg U, Abarenkov K (2019) The UNITE database for molecular
541 identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucl Acids Res 47:D259-
542 D264
543 Oberwinkler F, Riess K, Bauer R, Kirschner R, Garnica S (2013) Taxonomic re-evaluation of the
544 *Ceratobasidium-Rhizoctonia* complex and *Rhizoctonia butinii*, a new species attacking spruce. Mycol Progr
545 12:763-776
546 Oguz F, Koehler W (2016) URL decay at year 20: A research note. Journal of the Association for Inform Sci
547 Technol 67:477-479
548 Pavlova NA, Sokornova SV (2018) Effect of drying on viability of different ages mycelium of *Stagonospora*
549 *cirsii*. Plant Protect News 4:67-69
550 Peay KG, Kennedy PG, Talbot JM (2016) Dimensions of biodiversity in the Earth mycobiome. Nat Rev
551 Microbiol 14:434-447
552 Põlme S, Bahram M, Jacquemyn H, Kennedy P, Kohout P, Moora M, Oja J, Öpik M, Pecoraro L, Tedersoo L
553 (2018) Host preference and network properties in biotrophic plant–fungal associations. New Phytol 217:1230–
554 1239

555 Pöggeler S, Wöstemeyer J (2011) The Mycota XIV. Evolution of Fungi and Fungal-Like Organisms. Springer,
556 Heidelberg
557 Promputtha I, Jeewon R, Lumyong S, McKenzie EHC, Hyde KD (2005) Ribosomal DNA fingerprinting in the
558 identification of non sporulating endophytes from Magnolia liliifera (Magnoliaceae). Fung Divers 20:167-186
559 Promputtha I, Lumyong S, Dhanasekaran V, McKenzie EHC, Hyde KD, Jeewon R (2007) A phylogenetic
560 evaluation of whether endophytes become saprotrophs at host senescence. Microb Ecol 53:579-590
561 Rambold G, Agerer R (1997) DEEMY--the concept of a characterisation and determination system for
562 ectomycorrhizae. Mycorrhiza 7:113-116
563 Rambold G, Zedda L, Coyle JR, Peršoh D, Köhler T, Triebel D (2016) Geographic heat maps of lichen traits
564 derived by combining LIAS light description and GBIF occurrence data, provided on a new platform. Biodiv
565 Conserv 25:2743-2751
566 Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, Levasseur A, Lombard V, Morin E, Otillar
567 R, Lindquist EA (2014) Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white
568 rot/brown rot paradigm for wood decay fungi. Proc Natl Acad Sci USA 111:9923-9928
569 Rodriguez RJ, White JF, Arnold AE, Redman RS (2009) Fungal endophytes: diversity and functional roles.
570 New Phytol 182:314-330
571 Ryan ET, Hill DR, Solomon T, Endy TP, Aronson N (2019) Hunter's Tropical Medicine and Emerging
572 Infectious Diseases E-Book. Elsevier, Canda
573 Schoch CL, Robbertse B, Robert V, Vu D, Cardinali G, Irinyi L, Kõljalg U, Tedersoo L, Federnse S (2014)
574 Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi.
575 Database 2014:1-21
576 Selosse MA, Schneider- Maunoury L, Martos F (2018) Time to re- think Fung Ecol? Fungal ecological niches
577 are often prejudged. New Phytol 217:968-972
578 Shearer CA and Raja HA (2010) Freshwater Ascomycetes Database: http://fungi.life.illinois.edu/ (last accessed
579 01.08.2020)
580 Smith SE, Read DJ (2008) Mycorrhizal Symbiosis, 3rd edn. 787 pp. Academic Press, London, UK
581 Soonvald L, Loit K, Runno-Paurson E, Astover A, Tedersoo L (2019) The role of long-term mineral and organic
582 fertilisation treatment in changing pathogen and symbiont community composition in soil. Appl Soil Ecol
583 141:45-53
584 Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D (2006) Eukaryotic microbes, species recognition
585 and the geographic limits of species: examples from the kingdom Fungi. Phil Trans R Soc B 361:1947-1963
586 Tedersoo L, Abarenkov K, Nilsson RH, Schüβler A, Grelet G-A, Kohout P, Oja J, Bonito GM, Veldre V, Jairus
587 T, Ryberg M, Larsson K-H, Kõljalg U (2011) Tidying up International Nucleotide Sequence Databases:
588 ecological, geographical and sequence quality annotation of ITS sequences of mycorrhizal fungi. PLoS ONE
589 6:e24940
590 Tedersoo L, Anslan S, Bahram M, Drenkhan R, Pritsch K, Buegger F, Padari A, Hagh-Doust N, Mikryukov V,
591 Kõljalg U, Abarenkov K (2020) Regional-scale in-depth analysis of soil fungal diversity reveals strong pH and
592 plant species effects in Northern Europe. Front Microbiol 11:1953
593 Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Villarreal-Ruiz L, Vasco-Palacios A,
594 Quang Thu P, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Ratkowsky D, Pritsch K, Riit T, Põldmaa K,

595 Piepenbring M, Phosri C, Peterson M, Parts K, Pärtel K, Otsing E, Nouhra E, Njouonkou AL, Nilsson RH,
596 Morgado LN, Mayor J, May TW, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo L, Greslebin
597 A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkhan R, Dearnaley J, De Kesel A, Dang T, Chen X,
598 Buegger F, Brearley FQ, Bonito G, Anslan S, Abell S, Abarenkov K (2014) Global diversity and geography of
599 soil fungi. Science 346:1078
600 Tedersoo L, Bahram M (2019) Mycorrhizal types differ in ecophysiology and alter plant nutrition and soil
601 processes. Biol Rev 94:1857–1880
602 Tedersoo L, Sánchez-Ramírez S, Kõljalg U, Bahram M, Döring M, Schigel D, May T, Ryberg M, Abarenkov K
603 (2018a) High-level classification of the Fungi and a tool for evolutionary ecological analyses. Fung Divers
604 90:135–159
605 Tedersoo L, Smith ME (2013) Lineages of ectomycorrhizal fungi revisited: foraging strategies and novel
606 lineages revealed by sequences from belowground. Fung Biol Rev 27:83-99
607 Tedersoo L, Smith ME (2017) Ectomycorrhizal fungal lineages: detection of four new groups and notes on
608 consistent recognition of ectomycorrhizal taxa in high-throughput sequencing studies. Ecol Stud 230:125-142
609 Tedersoo L, Tooming-Klunderud A, Anslan S (2018b) PacBio metabarcoding of fungi and other eukaryotes:
610 biases and perspectives. New Phytol 217:1370-1385
611 Tisthammer KH, Cobian GM, Amend AS (2016) Global biogeography of marine fungi is shaped by the
612 environment. Fung Ecol 19:39-46
613 Triebel D, Peršoh D, Nash TH III, Zedda L, Rambold G (2007) LIAS – an interactive database system for
614 structured descriptive data of Ascomycetes. In: Currey GB, Humphries CJ (Eds) Biodiversity Databases.
615 Techniques, Politics, and Applications. Syst Assoc Spec 73:99–110
616 Větrovský T, Morais D, Kohout P, Lepinay C, Algora C, Hollá SA, Bahnmann BD, Bílohnědá K, Brabcová V,
617 D'Alò F, Human ZR, Jomura M, Kolařík M, Kvasničková J, Lladó S, López-Mondéjar R, Martinović T,
618 Mašínová T, Meszárošová L, Michalčíková L, Michalová T, Mundra S, Navrátilová D, Odriozola I, Piché-
619 Choquette S, Štursová M, Švec K, Tláskal V, Urbanová M, Vlk L, Voříšková J, Žifčáková L, Baldrian P
620 (2020). GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing
621 metabarcoding studies. Sci Data 7:1-14
622 Weete JD, Abril M, Blackwell M (2010) Phylogenetic distribution of fungal sterols. PLoS One 5:e10899
623 Zanne AE, Abarenkov K, Afkhami ME, Aguilar-Trigueros CA, Bates S, Bhatnagar JM, Busby PE, Christian N,
624 Cornwell W, Crowther TW, Moreno HF (2020) Fungal functional ecology: Bringing a trait-based approach to
625 plant-associated fungi. Biol Rev 95:409-433
626

627 Fig. 1. Distribution of fungal genera by primary lifestyle in each fungal phyla as well as Stramenopiles. Included
628 are primary lifestyles that exceed 0.5% of fungal genera.

629 Fig. 2. Distribution of annotated fungal sequences by DNA source. For 'DNA isolation source', trait states
630 exceeding 1% abundance are presented; for 'Culture source', traits states exceeding 2% abundance are
631 presented.

Fig. 3. The most common trait states of the species hypothesis with the largest number of sequences (SH1688425.08FU), roughly corresponding to a single biological species, *Alternaria eichhorniae*.

Fig. S1. Trait distributions of fungal genera in different fungal phyla.

Fig. S2. Trait distributions of Stramenopila genera in different Stramenopila phyla.

Fig. S3. Distribution of the ten most common fungal guilds among annotated sequences.

Table S1. Traits of genera.

Table S2. Traits of sequences.

Table S3. Traits of species hypothesis.

Table S4. Example dataset for genus-level annotation using the *vlookup* function in Excel.

Table S5. Comparison of workflows and outputs conducted in FunTraits and FUNGuild.

Supplementary item 1. List of trait states for genera and sequences.

Supplementary item 2. Instructions for annotators of fungal ITS sequences.

Fig. 1

Click here to access/download;colour figure;Fig. 1.eps

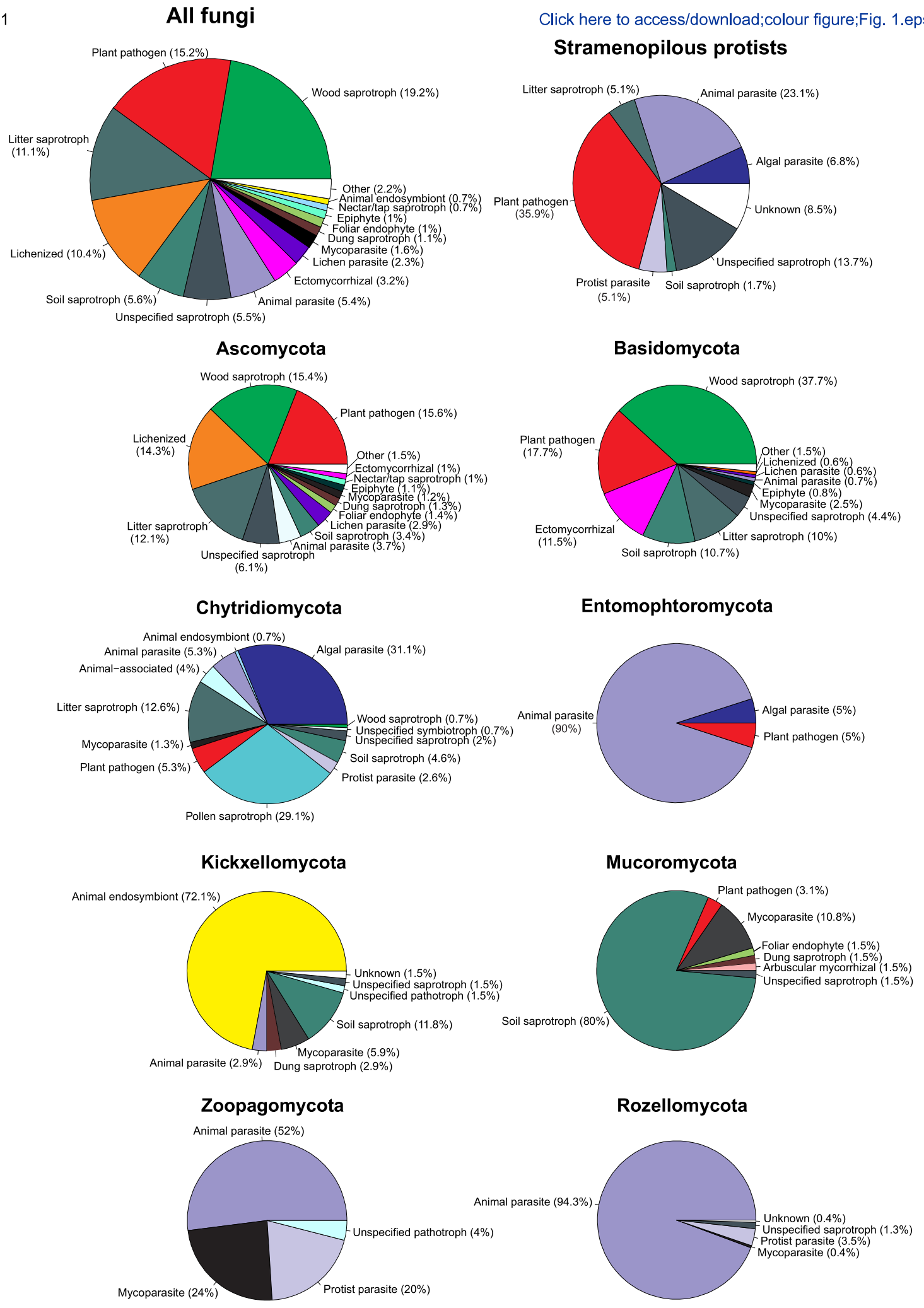
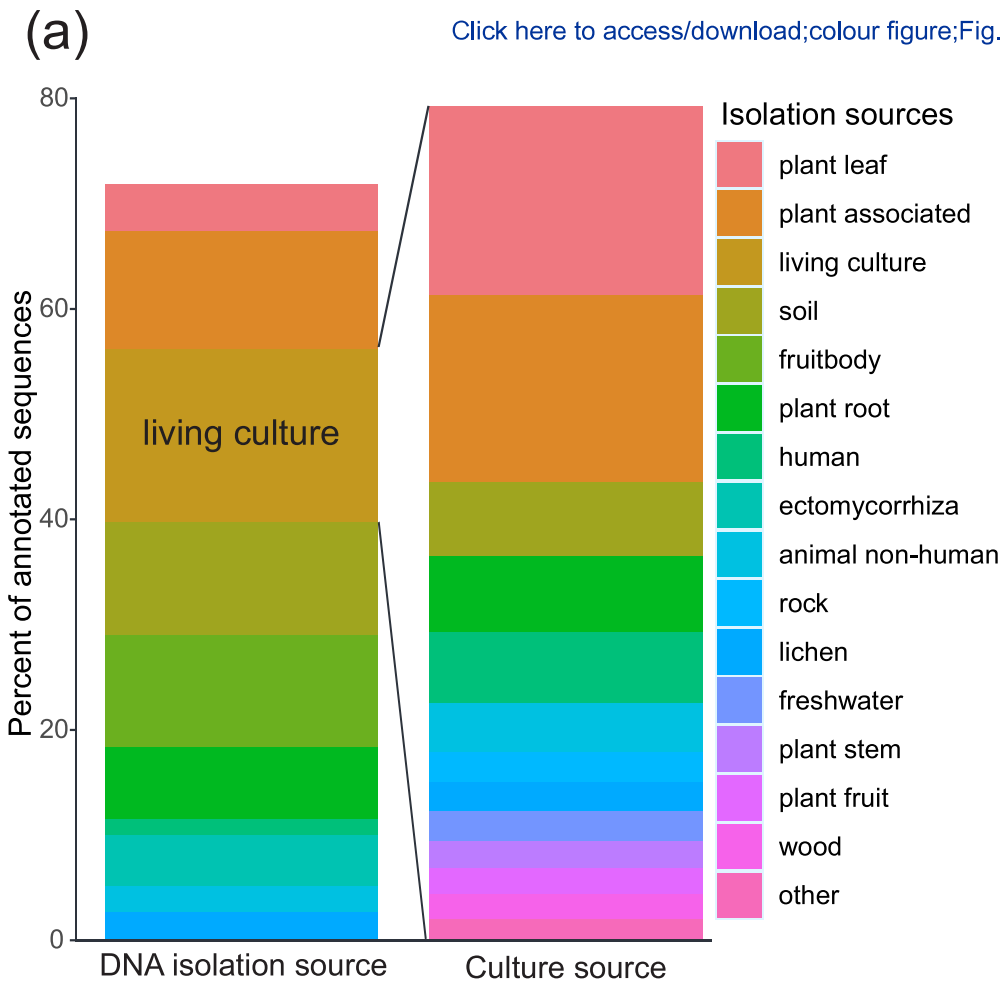


Fig. 2

Click here to access/download;colour figure;Fig. 2.eps



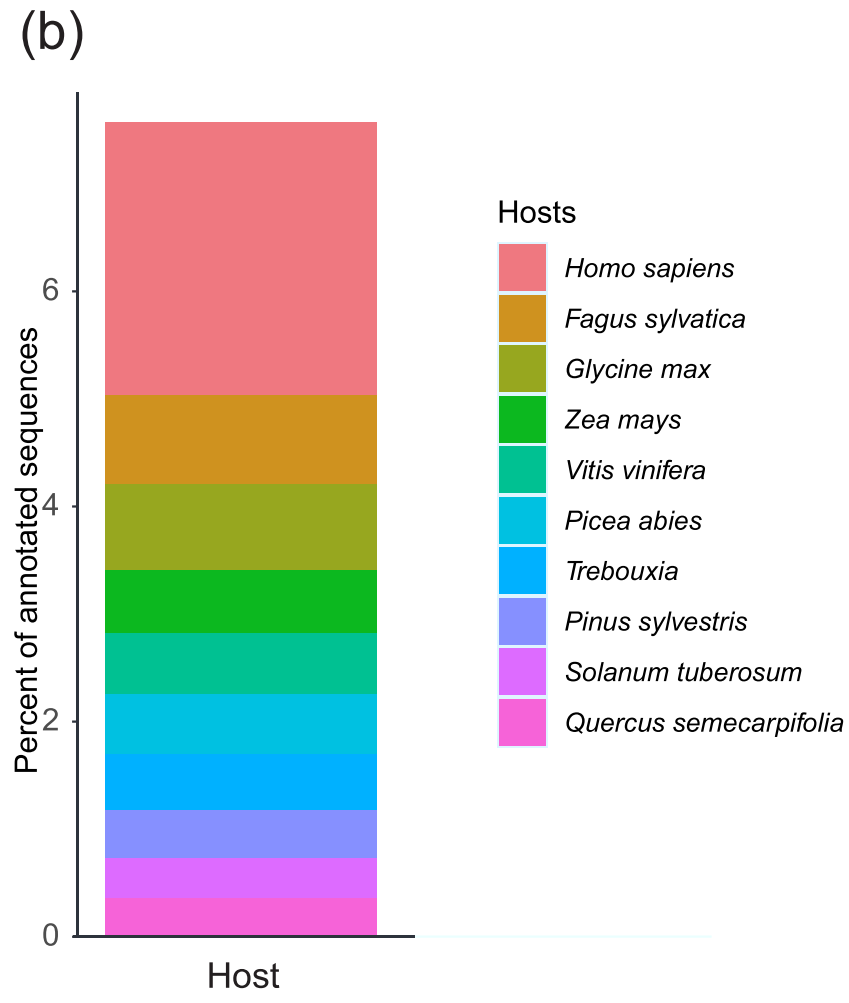


Fig. 3

Click here to access/download;colour figure;Fig. 3.eps