



OPEN

Citizen science helps in the study of fungal diversity in New Jersey

DATA DESCRIPTOR

Maria Shumskaya¹✉, Nina Filippova², Laura Lorentzen¹, Shazneka Blue¹, Carrie Andrew³ & Nicholas S. Lorusso^{1,4}

The history of fungal diversity of the Northeastern United States is currently fragmentary and restricted to particular functional groups or limited geospatial scales. Here, we describe a unique by its size, lifespan and data originators dataset, to improve our understanding of species occurrence and distribution across the state and time. Between the years 2007 to 2019, over 30 parks and nature preserves were sampled during forays conducted by members of the New Jersey Mycological Association (USA), a nonprofit organization of fungi enthusiasts. The dataset contains over 400 000 occurrences of over 1400 species across the state, made up mostly of the phylum Basidiomycota (89%) and Ascomycota (11%), with most observations resolved at the species level (>99%). The database is georeferenced and openly accessible through the Global Biodiversity Information Facility (GBIF) repository. This dataset marks a productive endeavor to contribute to our knowledge of the biodiversity of fungi in the Northeastern United States leveraging citizen science to better resolve biodiversity of this critical and understudied kingdom.

Background & Summary

Fungi are highly diverse and crucial for a wide variety of ecosystem services. They are one of the largest groups of decomposers, playing an essential role in nutrient cycling as saprotrophs or mycorrhiza, and facilitating ecosystem feedback to climate changes^{1–4}. Fungi are a species-rich taxon of approximately 2.2–3.8 million species⁵, lesser in number to only compared with terrestrial arthropods (over 7 million species⁶). The global inventory of fungal species presents a greater challenge compared to other taxonomic groups, due to a temporal nature of bodies of the most fungi. For example, plants (ca 374 000 species⁷) have a lower estimated species number, but are much better studied and documented in a biodiversity context. While plants do play critical roles in global carbon cycling, justifying historical focus on that group, fungi play an equally important role as decomposers and nutrient cyclers. Functional groups such as the mycorrhizae are key players in regulation of carbon dioxide⁸, nitrogen and phosphorus cycling^{9,10}, with up to 80% of nutrient cycling being provided via fungal symbionts. Given that much of the nutrient cycling associated with gross primary production as well as climate change (mediated through carbon dioxide consumption by plants) may be so strongly tied to fungi, understanding their diversity is critically important, making the bias against fungi a major gap in our understanding of taxa relevant to these cycles. Occasional surveys of fungal diversity based on fruiting bodies have been conducted for limited geographic scales¹¹, however, more often fungi are included in larger biodiversity surveys but are disproportionately under described in resulting datasets^{12,13}. While some large-scale studies of certain important fungal groups like mycorrhizae or parasitic species have been conducted^{14,15}, broader evaluation of fungal biodiversity at larger spatial scales is still limited. Publicly available datasets^{11,12,16–19} of fungi provide critical data for specialists studying biodiversity, ecology, environmental science, and mycology but more data is required to access the full extent of global fungal diversity. With recent developments in DNA sequencing technologies allowing detection and description of fungal species based on DNA evidence alone, the number of fungal species is expected to rapidly grow²⁰.

The limited detectability, attraction for research, and identification of fungi by both professionals and amateurs are constantly expanding. This rapidly growing inclusion of fungi into our overall study of biodiversity will only continue to expand existing DNA reference libraries²¹. Citizen science, too, takes new shapes as not

¹Department of Biology, Kean University, 1000 Morris Ave, Union, NJ, 07083, USA. ²Yugra State University, Chekhova str., 16, Khanty-Mansiysk, 628012, Russia. ³Oberlin College & Conservatory, Biology Department, 119 Woodland Street, Oberlin, Ohio, 44074, USA. ⁴Department of Natural Sciences, University of North Texas at Dallas, 7300 University Hills Blvd, Dallas, TX, 75241, USA. ✉e-mail: mshumaska@kean.edu



Fig. 1 A map of the state of New Jersey, USA, where occurrence data were collected. Colored symbols: sub regions showing geographic average of the sampled sites. CE – Central East, COAST – Coastal, CW – Central West, N – North, NE – North East, NW – North West, SNC – South Non-Coastal, SW – South West, W - West.

only local communities, but entire mycological congresses²² use approaches such as BioBlitz to complement traditional surveys.

At present, nature conservation efforts for fungi as are still developing in the United States^{23,24}, as prior survey initiatives have been largely scattered and maintained by voluntary efforts of resident regional mycologists (amateur or professional). Due to the fragmented nature of previous observations, there is a need to resolve fungal species diversity at regional scales in the United States, especially in highly populated areas comprised of multiple habitat and ecosystem types influenced by the legacy of citizen science initiatives^{5,25,26}. For states comprised of a variety of different ecosystem types, such as New Jersey, this presents a major gap in knowledge given the degree of geographic and ecological diversity seen state-wide. While some efforts have been made to characterize fungal diversity for New Jersey, these attempts have largely been restricted to smaller geographic scales or specific taxonomic groups of interest such as lichens^{27,28} or parasitic species²⁹. Despite these historical limitations, there is a pressing need to better characterize fungal biodiversity at larger spatial scales with recent work at global scales emerging for specific fungal guilds^{30–34}.

The challenge presented by the lack of sufficient fungal diversity data can be resolved, to some extent, using citizen science. With the recent development of digital technologies, citizen science has been successful in contributing to fundamental research^{35–37}. Several online platforms collect photographic and written observations directly from citizens (e.g. mushroomobserver.org, iNaturalist.org or fundis.org). However, the observations made via these portals are mostly recent; the tremendous efforts of amateur groups who have been tracking fungal diversity before the digital era are often left closed to the global research community due to the lack of the proper data storage and sharing protocols and so it was not uncommon for the local organizations to keep records on a personal computer or in a hand-written format. This limitation in protocols for sharing data collected by citizen scientists presents one of the major opportunities to researchers of biodiversity at larger spatial scales and making these datasets openly available is critically important to better resolve global biodiversity.

Here, we describe a dataset consisting of fungal taxa for the state of New Jersey collected as part of citizen science forays in 32 parks and nature preserves throughout the state, separated into nine sub regions³⁸ (Fig. 1), Tabel 1.

Data included in this dataset were collected between the year 2007 and 2019, by volunteers of the New Jersey Mycological Association (NJMA, www.njmyco.org) as part of their organization's yearly sampling forays. Established in 1971 as the Lakeland Mycology Club, NJMA is now non-profit organization with over 800 members motivated by their interest in fungi, the only organization of its kind in the state of New Jersey. NJMA has amassed a wealth of citizen science data through decades of sampling events. Importantly, NJMA maintains an active herbarium of approximately 3000 vouchered specimens stored at Rutgers University in New Brunswick;

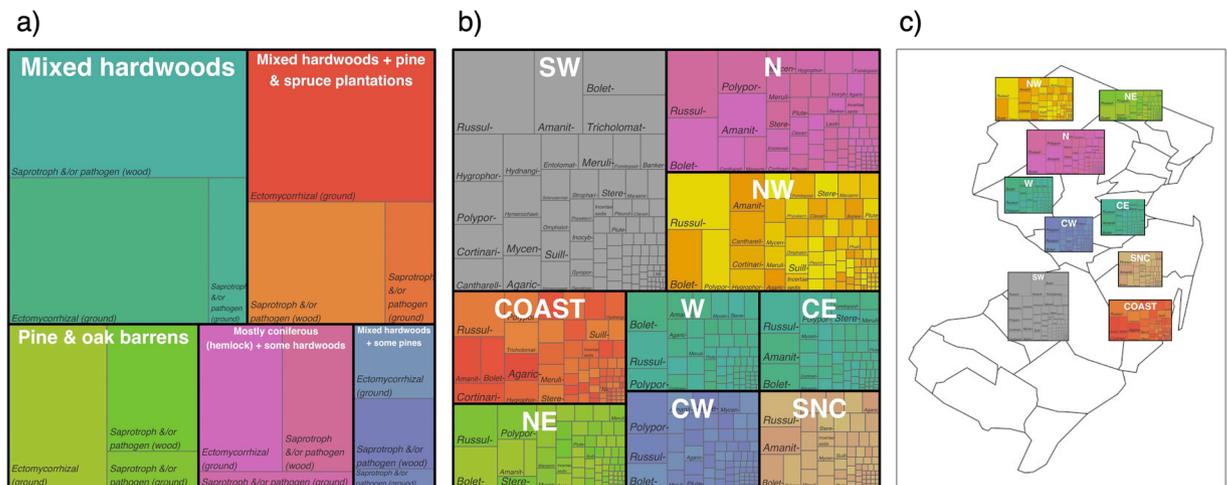


Fig. 2 Trends for fungal taxa in the dataset. **(a)** differences in the relative proportion of common ecological guilds seen in different sample sites categorized by primary forest type. **(b)** regions sampled across the state and families of fungi found within these regions. **(c)** regions sampled across the state plotted onto a map of New Jersey and families of fungi found within these regions. In all plots of the boxes is proportional to the number of observations for those habitat types, regions, guilds, or functional groups. CE – Central East, COAST – Coastal, CW – Central West, N – North, NE – North East, NW – North West, SNC – South Non-Coastal, SW – South West, W – West.

however, this repository is currently not a part of the Chrysler Herbarium (CHRB) at Rutgers. Here, we showcase how large quantities of data collected across a variety of habitats and locations over a span of 12 years by volunteers has contributed to scientific knowledge in a cost-effective and data rich manner. Our interest was to prepare the collected data in a standardized format and make it open access, to increase the applicability to fungal biodiversity research. The resultant dataset is also meant to raise interest among citizen science and scientists to increase the amount of accessible data on the distribution of species^{37,39}. Given that the North American Mycological Association (NAMA, <https://namyco.org/clubs.php>) has records of over 90 similar groups across 37 states this type of citizen science driven data collection has the potential to exponentially increase our knowledge of fungal taxa across the United States.

The dataset presented here³⁸ highlights the taxonomic diversity for the state of New Jersey from 210 surveys (corresponds to 210 records in the event table), with 400 260 records in the occurrence table. Overall, 1906 taxa with presence/absence information for each survey are published. In total, 96% of the records in the occurrence table are absence data. The taxonomic structure is presented by 2 kingdoms (Fungi and Protista for slime molds), 5 phyla, 20 classes, 58 orders, 162 families, 516 genera and 1483 species. Some species names that were assigned to the observations earlier in the data collection are now outdated, but are kept in the dataset as synonyms or under-identified taxa. Together with these records, the total number of taxa sums up to 1850. Taxa varied both by environment type based on primary forest composition (Fig. 2a) and by region (Fig. 2b), although the regional effect may also be influenced by a sampling month.

The dataset is supplemented with openly available environmental variables of interest (average region temperature; maximum region temperature; average region humidity due point; average region precipitation; average region wind; retrieved from the United States Geological Survey <https://www.usgs.gov> and National Weather Service <https://www.weather.gov> databases) for each sampling location in order to provide insight into factors contributing to changes in the distribution of taxa. This information is available for each sampling event (survey) in the dataset at GBIF.org.

Despite the variable nature of collection across this period (such as different frequency of visiting of the same sampling sites, Table 2), this dataset presents an opportunity for researchers and citizen scientists interested in fungal biodiversity of the Northeastern United States. The dataset describes relative abundances for common taxa over time. Their trophic types are presented on Fig. 3. To expand our understanding of how this dataset compares to other similar attempts to capture fungal diversity across geographic scales, we compared our guild data to other Agaricomycetes datasets published at GBIF.org for New Jersey, the United States, and globally (Fig. 4). Despite some variability in specific species present across these spatial scales, we found similar proportions of the guild types (Fig. 4). Together, this suggests that our dataset captures information similar to other datasets of this type when considering functional roles of fungi, while adding to our knowledge of region-specific introduced or newly observed species as biodiversity changes globally (Tables 3, 4).

We suggest that this dataset, and the associated collection methods used by the New Jersey Mycological Association, could be used as a model for a systemic approach for evaluating fungal diversity across the United States. Through increasing open and digital access to fungal data, we expect that the presented dataset contributes to more complete documentation of life on Earth beyond charismatic taxa.

Sampling Locality	Latitude	Longitude	Generalized Habitat	Region
Cheesequake State Park	40.43615	-74.2653	Mixed hardwoods	CE
Helmetta bog	40.38147	-74.4312	Mixed hardwoods	CE
Holmdel County Park	40.37041	-74.1843	Mixed hardwoods	CE
Thompson Park	40.33403	-74.4403	Mixed hardwoods	CE
Cattus Island Park	39.97802	-74.138	Pine & oak barrens	COAST
Jakes Branch County Park	39.92915	-74.2096	Pine & oak barrens	COAST
Wells Mills County Park	39.79729	-74.278	Pine & oak barrens	COAST
Herrontown Woods	40.37568	-74.6397	Mixed hardwoods	CW
Princeton Institute Woods	40.33075	-74.6594	Mixed hardwoods	CW
Washington Crossing State Park	40.29818	-74.8672	Mixed hardwoods + pine & spruce plantations	CW
Meadowood Park	40.79138	-74.6424	Mixed hardwoods	N
Schiff Nature Preserve	40.75003	-74.6298	Mixed hardwoods	N
Stephens State Park	40.86891	-74.8102	Mixed hardwoods	N
New Weis Center for Education, Arts, and Recreation	41.06979	-74.3211	Mixed woods	NE
Wawayanda State Park	41.19632	-74.392	Mixed woods (hemlock stands, mixed hardwoods)	NE
Pocono Environmental Education Center	41.17124	-74.9144	Mostly coniferous (hemlock) + some hardwoods	NW
Stokes State Forest – Kittle Field	41.18439	-74.7973	Mostly coniferous (hemlock) + some hardwoods	NW
Stokes State Forest – NJ School of Conservation	41.22571	-74.7518	Mostly coniferous (hemlock) + some hardwoods	NW
Stokes State Forest, Lake Ocquittunk	41.2294	-74.7652	Mostly coniferous (hemlock) + some hardwoods	NW
Belleplain State Forest	39.24875	-74.8415	Pine barrens	SNC
Estell Manor Park	39.3983	-74.7454	Pine barrens	SNC
Forest Resource Education Center	40.09304	-74.3231	Pine & oak barrens	SNC
Manasquan Reservoir Environmental Center	40.17787	-74.2221	Mixed woods/oak barrens	SNC
Ocean County Park	40.09062	-74.1903	Pine & oak barrens	SNC
Brendan T. Byrne Park and State Forest	39.87479	-74.5474	Pine barrens	SW
Chestnut Branch Park	39.77129	-75.1672	Mixed hardwoods + some pines (?)	SW
Franklin Parker Preserve	40.24178	-74.9678	Pine & oak barrens	SW
Rancocas State Park	40.0041	-74.821	Pine & oak barrens	SW
Deer Path Park	40.55654	-74.8434	Mixed hardwoods	W
Hoffman County Park	40.62797	-74.9949	Mixed hardwoods (mostly oak)	W
Horseshoe Bend Park	40.50603	-75.0415	Mixed hardwoods	W
Teetertown Ravine Nature Preserve and Crystal Springs	40.75654	-74.8518	Mixed hardwoods	W

Table 1. Localities sampled as part of NJMA yearly forays with representative coordinates, generalized habitat, and user defined regions within the state. CE – Central East, COAST – Coastal, CW – Central West, N – North, NE – North East, NW – North West, SNC – South Non-Coastal, SW – South West, W – West.

Methods

Geographic coverage. The state of New Jersey is a spatially small (22 610 km²) yet densely populated (8.885 million) member of the mid-Atlantic states in the Northeastern United States. Comprised of a number of ecosystem types, including coastal regions to the east, deciduous forests in the central and northern parts of the state, and pine forests/barrens in the south, New Jersey presents a diversity of ecological community types. The state ranges between sea level and 550 meters and is flat with an average elevation of only 69 meters. The state is also marked by a high population density and degree of human use, with roughly a third of total land for the state currently developed. The state saw continuous development during the sampling duration for the dataset with an increase of more than 360 000 acres of land being developed between the years 1986 and 2015 (NJDEP, 2020, <https://www.nj.gov/dep/dsr/trends/>). The climate follows four regular seasons typical of the Northeastern United States with most of the state characterized as humid subtropical. Average rainfall ranges between 1016–1295 mm. Average yearly temperature is roughly 13.3 °C. Both average temperature and precipitation have been recently increasing, consistent with climate change.

Study extent. Between 2007 and 2019, the NJMA conducted yearly surveys, where members would visit parks across the state at approximately the same time each year to conduct sampling forays. Most sampling sites were parks or nature preserves (Table 1) selected due to their probability of having representative fungal biodiversity for that part of the state. Samples for the foray locations were collected for identification from within the boundaries of the sampling sites.

Sampling locality by region	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Central East	•	•	•	•	•		•	•	•				
Cheesequake State Park												•	
Helmetta bog						•	•	•	•				
Holmdel County Park									•	•	•	•	•
Thompson Park													
Coastal													
Cattus Island Park		•	•	•	•	•				•		•	•
Jakes Branch County Park	•	•	•	•	•	•	•	•	•	•	•	•	•
Wells Mills County Park	•	•		•	•	•	•	•					
Central West													
Herrontown Woods	•	•	•										
Princeton Institute Woods	•	•	•	•		•	•	•	•	•	•	•	•
Washington Crossing State Park	•	•		•	•	•	•	•					
North													
Meadowood Park	•	•	•	•	•	•	•	•	•	•	•		•
Schiff Nature Preserve	•	•	•	•	•	•	•	•				•	
Stephens State Park	•	•	•	•		•	•		•	•	•	•	•
North East													
New Weis Center for Education, Arts, and Recreation											•	•	
Wawayanda State Park			•	•	•	•	•	•	•	•	•	•	•
North West													
Pocono Environmental Education Center	•							•					
Stokes State Forest-Kittle Field -	•	•	•	•	•	•	•	•	•	•	•	•	•
Stokes State Forest-NJ School of Conservation												•	•
Stokes State Forest, Lake Ocquittunk			•	•	•		•	•	•	•	•		•
South Non-Coastal													
Belleplain State Forest									•	•	•	•	•
Estell Manor Park												•	•
Forest Resource Education Center										•	•	•	•
Manasquan Reservoir Environmental Center	•	•	•	•	•	•	•	•	•				
Ocean County Park											•		
South West													
Brendan T. Byrne Park and State Forest	•	•	•	•	•		•	•	•	•	•	•	
Chestnut Branch Park													•
Franklin Parker Preserve			•	•	•	•		•	•	•	•		
Rancocas State Park	•	•	•	•	•	•				•	•		
West													
Deer Path Park			•										
Hoffman County Park	•	•	•	•	•	•	•	•	•				
Horseshoe Bend Park								•				•	
Teetertown Ravine Nature Preserve and Crystal Springs											•	•	•

Table 2. Sampling localities organized by both the region and the year of sampling (2007–2019).

Sampling sites were distributed across the entire state of New Jersey (Fig. 1) and were visited with some temporal variability within each year. Some foray locations within the dataset were unique to only part of the 12-year sampling window, while other foray locations were sampled consistently across the entire sampling period (e.g. Pocono Environmental Education Center vs Wawayanda State Park, Table 2). Sites were sampled between May and November of each year with citizen scientists sampling one site per day. Sites were normally sampled during the same month across years, though some variation in sampling time did occur. To better describe similar

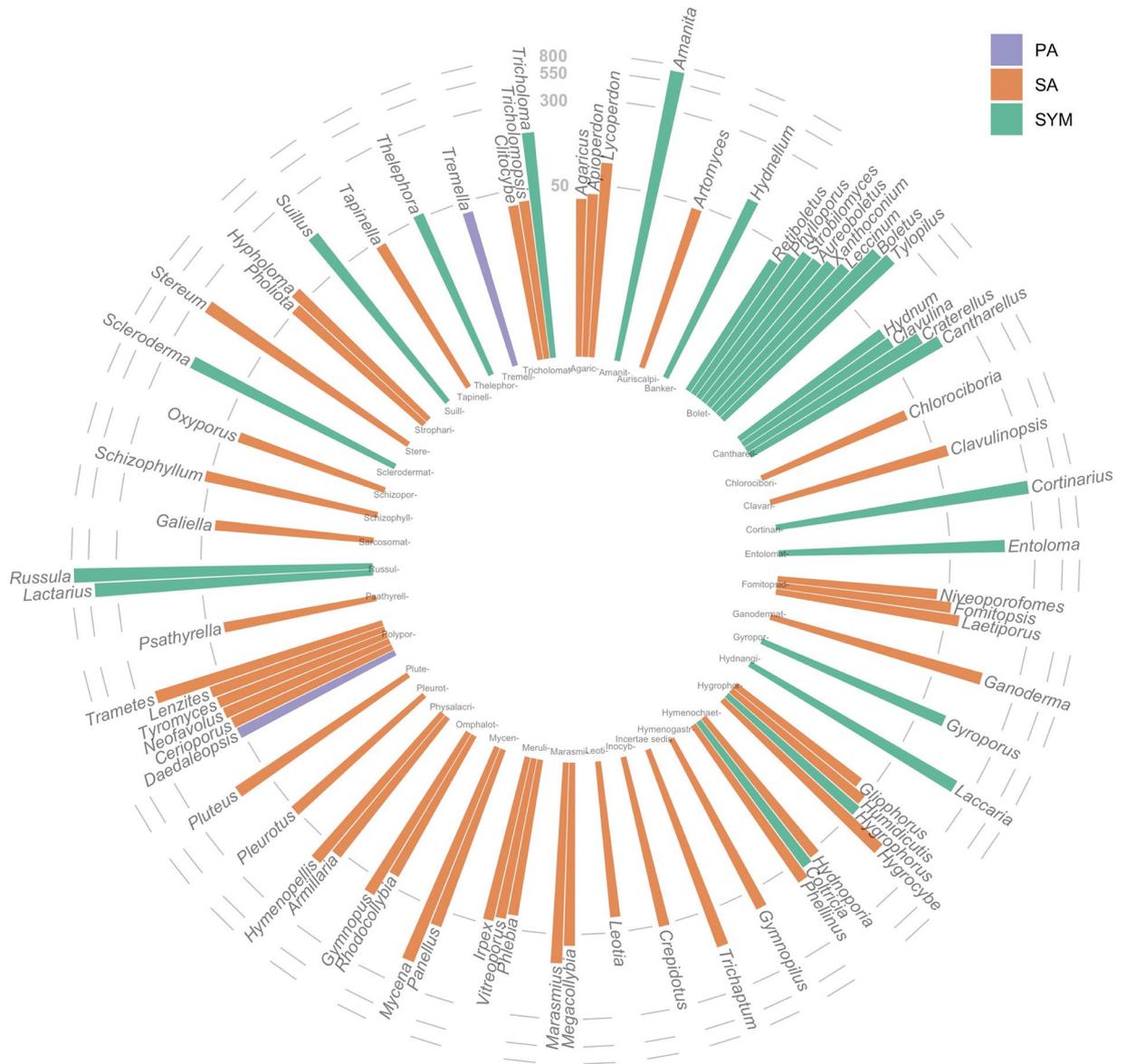


Fig. 3 Most commonly observed taxa across all sampling sites and years shown as unscaled numbers of observations on a log₁₀ scale. Colored by trophic type (sensu FunGuild,⁵¹ PA - pathotroph, SA - saprotroph, SYM - ectomycorrhizal symbiotroph) and sorted by the most common fungal families, with family names on the inside of the ring. Tips of each bar: names of commonly observed genera.

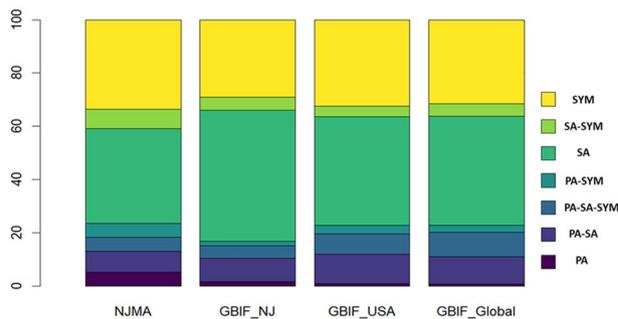


Fig. 4 Proportion of seven guild types (sensu FunGuild⁵¹) in Agaricomycetes for the evaluated datasets. SYM - symbiotroph, SA- saprotroph, PA- pathotroph. NJMA: from NJMA dataset⁴⁰, GBIF_NJ: records retrieved from GBIF.org for New Jersey⁴¹, GBIF_USA: records retrieved from GBIF.org for the United States⁴², GBIF_global: global records retrieved from GBIF.org⁴³.

Dataset name (shared/unique species)	NJMA	GBIF NJ	GBIF USA	GBIF Global
NJMA	1248 species	NJMA: 691 unique GBIF NJ: 605 unique Shared: 557	NJMA: 32 unique GBIF USA: 7636 unique Shared: 1216	NJMA: 4 unique GBIF Global: 28349 unique Shared: 1244
GBIF NJ		1162 species	GBIF NJ: 0 unique GBIF USA: 7690 unique Shared: 1162	GBIF NJ: 0 unique GBIF Global: 28431 unique Shared: 1162
GBIF USA			8852 species	GBIF USA: 0 unique GBIF USA: 20741 unique Shared: 8852
GBIF Global				29592 species

Table 3. Species counts for Agaricomycetes datasets published at GBIF.org. NJMA: from the dataset of this study⁴⁰. GBIF NJ: the dataset of preserved specimens for New Jersey region⁴¹. GBIF USA: the dataset of preserved specimens for the USA⁴². GBIF Global: global records of preserved specimens⁴³. Total species counts are presented on the diagonal. Unique or shared species numbers are presented above the diagonal.

Dataset	Number of observations	Percentage of observations from total number of global observations
Artportalen (Swedish Species Observation System)	1661121	24
Observation.org, Nature data from around the World	747761	11
Danish Mycological Society, fungal records database	518310	7
Norwegian Species Observation Service	494795	7
iNaturalist Research-grade Observations	426203	6
Fungi of parks, forests and reserves of New Jersey (2007–2019)	400260	6
Swiss National Fungi Databank	271159	4
Österreichische Mykologische Gesellschaft - Austrian Mycological Society	163957	2
BLS Lichen Database: England 1650–2016	162753	2
BLS Lichen Database: Scotland 1700–2016	119995	2
Others	2090943	30

Table 4. Number of occurrences per dataset published in GBIF.org. Data were searched globally (using filters: kingdom - “fungi”, basis of record - “human observation”, occurrence status - “everything”, and year - “2007–2019”). Total number of observations found: 6 194 448.

regions across the state, we assigned regional identifications to each foray sampling locality based on habitat similarity and available climatological data. The sites sampled were selected to provide a measure of the fungal biodiversity within different ecosystems types representative of the state. Forays stopped in 2020 due to the COVID-19 pandemic and data after 2019 were not included due to changes in sampling activity.

Data acquisition. The established sampling foray method has been practiced by NJMA for the past 30 years. Sampling forays were conducted for two hours at each foray location with any member able to participate in collection. Some forays were made open to the public and participant numbers ranged from 5 to 30 people, with all participants starting from specific starting point. Within the two-hour foray period, samplers surveyed approximately a one-mile radius around the starting point and collected any visible sporocarps and returned them to foray leaders for identification. Hypogeous taxa were not explicitly sampled as part of these forays, and the focus was on macroscopic fruit bodies (with select observations of micro fungi). Sampled taxa were identified on site by foray leaders, and records were stored in the NJMA documentation archive with some samples stored in the herbarium. This combined strategy, using experts for identification and many participants for sample collection, effectively leveraged citizen science to make use of many samplers without formal scientific training in collection when a limited number of visiting taxonomic experts are available. The lists of observed species were recorded and saved as hand-written, PDF or Microsoft Office documents and stored at personal computers of NJMA members. A summary result of the forays was published in a PDF e-letter from the Association to its members and also shared on their website www.njmyco.org.

Datasets for different regional scales used in Fig. 4 and Table 4 were retrieved from GBIF.org and checked for accuracy to ensure species names matched across regional lists. A list of Agaricomycetes, a class highly represented in the NJMA dataset⁴⁰, was selected from the dataset and used to compare diversity at higher spatial scales. Checklists for preserved specimens within the Agaricomycetes were retrieved (filtered by basis of record - “preservedSpecimen”, occurrence status = “present”) for three regional scales: New Jersey⁴¹, the United States⁴², and global⁴³.

Taxonomic identification. Initial identification of taxa collected during forays was completed by foray leaders in the field using existing literature (listed in the GBIF repository⁴³) by assigning species names of the closest morphospecies.

Data digitalization and unification. The species lists were obtained from various NJMA members and converted from the existing format (Word, PDF, Excel) to Excel-based templates compatible with the EarthCape database (<https://earthcape.com/>⁴⁴), spell checking, formatting, association of data with information fields such as locality name or scientific taxon name was carefully performed. EarthCape allowed consolidation of locations into different user defined regions according to geographic location, habitat type, or climatic zone. The EarthCape database also confirmed consistent taxonomic synonymizing by comparison of user-assigned species identities against currently accepted taxonomic names of GBIF taxonomic backbone at GBIF.org, and allowed to convert the data to the GBIF format to prepare for the dataset publication.

Data Records

The dataset contains a description of whether a species (or in rare cases, a genus) were observed during a particular foray event. For all taxa observed across all forays, the presence or absence of that taxa is recorded in a particular foray and supplemented by the foray time and location, geographic data (coordinates, region, etc.), habitat type based on dominant hardwood in that location, and climatic variables (averages for temperature, precipitation, wind speed, dew point). Data on soil chemistry and geological variables were retrieved from United States Geological Survey (<https://www.usgs.gov>). Data for climate variables were retrieved from National Weather Service (www.weather.com) with representative collection stations identified and used for each region.

Our database is stored locally and is freely accessible through the GBIF (Global Biodiversity Information Facility) repository (www.gbif.org) under the <https://doi.org/10.15468/7scek438>. For each occurrence record there are 61 fields of information, recorded using terms of Darwin Core standard (DwC)⁴⁵ (<http://rs.tdwg.org/dwc/terms>). The database includes a supplemental data table that provides the climatological and geological data for each foray. These “Measurement or fact” extension table can be downloaded together with the source data. The dataset will be updated as new yearly forays occur to keep data consistent across forays. It is our intent that this data collection, made possible by the common interests among citizen science and scientists, continues to expand our knowledge of fungal distribution and biodiversity.

Technical Validation

The data was validated using standardized procedure for digitization, formatting, and content checking of the occurrence as in earlier studies⁴⁶. Integration and digitizing of data from various resources (electronic files, hand written documents etc.) was performed using EarthCape database software⁴⁴ built-in validation tools such as formatting and spelling checks, linked tables, alignment of nomenclature with the GBIF backbone⁴⁷, and synonymizing. To ensure the names and authors for all taxa observed, species names were confirmed using Index Fungorum (<http://www.indexfungorum.org>). Homotypic names were checked to refer to the accepted names. Next, the data was exported into three linked tables using Darwin Core standard⁴⁵: occurrence, event and measurementOrFact. The final data cleaning and processing was made using Linux command line scripts using bash and awk by R. Mesibov⁴⁸, and included structure, format and content of data. All data were checked again once prepared for publication via GBIF to validate the taxonomy, climate data, and the occurrence status. Our dataset was compared to similar records, as well as records at larger spatial scales, from GBIF.org (Table 3). We confirmed that all species from our dataset were captured at larger scales, with only several unique observations. When compared to the GBIF data available for New Jersey, our dataset shows 557 shared species. The difference with 691 species unique to GBIF records for New Jersey and 605 species unique to our dataset likely results from the study focus, regions sampled, and changes in the fungal composition of the state over the years. Our dataset also competes with some of the largest datasets for fungal biodiversity in GBIF, contributing significantly to the global data pool (6% of global fungal occurrences for the period of 2007–2019) (Table 4). Consistency of the taxonomic names was managed using GBIF Species API (<https://www.gbif.org/developer/species>) and the rgbif R package^{49,50}. Trophic type was assigned using R package `funguild`⁵¹.

Usage Notes

We suggest that the data within this dataset be used by researchers interested in evaluating large scale changes in biodiversity for fungi across space and time as well as researchers interested in studying the ranges of particular fungal taxa or guilds, for example, in remote sensing of mycorrhizal composition⁵². Beyond usage by formal researchers, we suggest our methods to be used by citizen science groups collaborating with universities and data repositories to make this data more accessible. The described method have proven to be efficient at leveraging citizen science records for fungal biodiversity and so we implore similar groups to consider reviewing our dataset with the included collection methods and planning their own forays using similar strategies. By leveraging shared collection methods across enthusiast societies with platforms for sharing data like GBIF and iNaturalist, we can greatly improve knowledge of fungal biodiversity across larger spatial scales.

Code availability

Figures were prepared using R Statistical Software (v4.1.2; R Core Team 2021)⁵³, packages `ggplot2`⁵⁴, `treemap`⁵⁵, `funguild`⁵¹ and `fungarium`⁵⁶. The dataset was shared via GBIF.org using Integrated Publishing Toolkit (IPT, www.gbif.org/ipt)⁵⁷. No original code was created to generate the dataset.

Received: 15 August 2022; Accepted: 19 December 2022;

Published online: 04 January 2023

References

- Martinez-Garcia, L. B., De Deyn, G. B., Pugnaire, F. I., Kothamasi, D. & van der Heijden, M. G. A. Symbiotic soil fungi enhance ecosystem resilience to climate change. *Glob. Chang. Biol.* **23**, 5228–5236 (2017).
- Averill, C. & Hawkes, C. V. Ectomycorrhizal fungi slow soil carbon cycling. *Ecol. Lett.* **19**, 937–947 (2016).
- Cairney, J. W. G. Extramatrical mycelia of ectomycorrhizal fungi as moderators of carbon dynamics in forest soil. *Soil Biol. Biochem.* **47**, 198–208 (2012).
- Milovic, M., Kebert, M. & Orlovic, S. How mycorrhizas can help forests to cope with ongoing climate change? *Sumar. List* **145**, 279–286 (2021).
- Hawksworth, D. L. & Luecking, R. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol. Spectr.* **5**, 5.4.10 (2017).
- Stork, N. E. How many species of insects and other terrestrial arthropods are there on Earth? *Annu. Rev. Entomol.* **63**, 31–45 (2018).
- Christenhusz, M. J. M. & Byng, J. W. The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201–217 (2016).
- Terrer, C., Vicca, S., Hungate, B. A., Phillips, R. P. & Prentice, I. C. Mycorrhizal association as a primary control of the CO₂ fertilization effect. *Science* **353**, 72–74 (2016).
- van der Heijden, M. G. A., Martin, F. M., Selosse, M. A. & Sanders, I. R. Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol.* **205**, 1406–1423 (2015).
- Braghiere, R. K. *et al.* Modeling global carbon costs of plant nitrogen and phosphorus acquisition. *J. Adv. Model. Earth Syst.* **14**, e2022MS003204 (2022).
- Jaouen, G. *et al.* Fungi of French Guiana gathered in a taxonomic, environmental and molecular dataset. *Sci. Data* **6**, 206 (2019).
- Beninde, J. *et al.* CaliPopGen: A genetic and life history database for the fauna and flora of California. *Sci. Data* **9**, 380 (2022).
- Gyeltshen, C. & Prasad, K. Biodiversity checklists for Bhutan. *Biodivers. Data J.* **10**, e83798 (2022).
- Steidinger, B. S. *et al.* Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* **569**, 404–408 (2019).
- Soudzilovskaia, N. A. *et al.* Global mycorrhizal plant distribution linked to terrestrial carbon stocks. *Nat. Commun.* **10**, 5077 (2019).
- Melo, C. D., Walker, C., Freitas, H., Machado, A. C. & Borges, P. A. V. Distribution of arbuscular mycorrhizal fungi (AMF) in Terceira and Sao Miguel Islands (Azores). *Biodivers. Data J.* **8**, e49759 (2020).
- Ordynets, A. *et al.* Aphyllophoroid fungi in insular woodlands of eastern Ukraine. *Biodivers. Data J.* **5**, e22426 (2017).
- Monteiro, M. *et al.* A database of the global distribution of alien macrofungi. *Biodivers. Data J.* **8**, e51459 (2020).
- Filippova, N. *et al.* Yugra State University Biological Collection (Khanty-Mansiysk, Russia): general and digitisation overview. *Biodivers. Data J.* **10**, e77669 (2022).
- Wu, B. *et al.* Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. *Mycology* **10**, 127–140 (2019).
- Nilsson, R. H. *et al.* The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* **47**, D259–D264 (2019).
- Gorczak, M. *et al.* 18th Congress of European Mycologists Bioblitz 2019 – naturalists contribute to the knowledge of mycobiota and lichenobiota of Białowieża Primeval Forest. *Acta Mycol.* **55**, 1–26 (2020).
- Goncalves, S. C., Haelewaters, D., Furci, G. & Mueller, G. M. Include all fungi in biodiversity goals. *Science* **373**, 403–403 (2021).
- Hochkirch, A. *et al.* A strategy for the next decade to address data deficiency in neglected biodiversity. *Conserv. Biol.* **35**, 502–509 (2021).
- Allen, E. B. *et al.* Patterns and regulation of mycorrhizal plant and fungal diversity. *Plant Soil* **170**, 47–62 (1995).
- Mueller, G. M. & Schmit, J. P. Fungal biodiversity: what do we know? What can we predict? *Biodivers. Conserv.* **16**, 1–5 (2007).
- Waters, D. P. & Lendemer, J. C. The lichens and allied fungi of Mercer County, New Jersey. *Opusc. Philolichenum* **18**, 17–51 (2019).
- Waters, D. P. & Lendemer, J. C. A revised checklist of the lichenized, lichenicolous and allied fungi of New Jersey. *Bartonia*, 1–62 (2019).
- Schwarze, C. A. *The parasitic fungi of New Jersey.* (New Jersey Agricultural Experiment Stations, 1917).
- Moose, R. A., Schigel, D., Kirby, L. J. & Shumskaya, M. Dead wood fungi in North America: an insight into research and conservation potential. *Nat. Conserv.* **32**, 1–17 (2019).
- Hibbett, D. S. *et al.* A higher-level phylogenetic classification of the Fungi. *Mycol. Res.* **111**, 509–547 (2007).
- Hibbett, D. The invisible dimension of fungal diversity. *Science* **351**, 1150–1151 (2016).
- James, T. Y., Stajich, J. E., Hittinger, C. T. & Rokas, A. Toward a Fully Resolved Fungal Tree of Life. *Annu. Rev. Microbiol.* **74**, 291–313 (2020).
- Braghiere, R. K. *et al.* Mycorrhizal distributions impact global patterns of carbon and nutrient cycling. *Geophys. Res. Lett.* **48**, e2021GL094514 (2021).
- Bonney, R. *et al.* Citizen science: A developing tool for expanding science knowledge and scientific literacy. *Bioscience* **59**, 977–984 (2009).
- Van Vliet, K. & Moore, C. Citizen science initiatives: engaging the public and demystifying science. *J. Microbiol. Biol. Educ.* **17**, 13–16 (2016).
- Feldman, M. J. *et al.* Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PLoS One* **16**, e0234587 (2021).
- Shumskaya, M. *et al.* Fungi of parks, forests and reserves of New Jersey (2007–2019). Version 1.4. Sampling event dataset. *Kean University* <https://doi.org/10.15468/7scek4> (2022).
- Heilmann-Clausen, J. *et al.* How citizen science boosted primary knowledge on fungal biodiversity in Denmark. *Biol. Conserv.* **237**, 366–372 (2019).
- GBIF.Org User. NJMA dataset. GBIF Occurrence Download. *GBIF* <https://doi.org/10.15468/dl.93232n> (2022).
- GBIF.Org User. New Jersey Agaricomycetes. GBIF Occurrence Download. Dataset. *GBIF* <https://doi.org/10.15468/dl.6j6382> (2022).
- GBIF.Org User. USA Agaricomycetes. GBIF Occurrence Download. *GBIF* <https://doi.org/10.15468/dl.ncukzy> (2022).
- GBIF.Org User. Global records Agaricomycetes. GBIF Occurrence Download. *GBIF* <https://doi.org/10.15468/dl.nk54e7> (2022).
- Meyke, E. When data management meets project management. *Biodivers. Inf. Sci. Stand.* **3**, e37224 (2019).
- Wieczorek, J. *et al.* Darwin Core: an evolving community-developed biodiversity data standard. *PLoS One* **7**, e29715 (2012).
- Pagad, S., Genovesi, P., Carnevali, L., Schigel, D. & McGeoch, M. A. Data Descriptor: introducing the global register of introduced and invasive species. *Sci. Data* **5**, 170102 (2018).
- Registry-Migration. *Gbif.Org*. GBIF Backbone Taxonomy. *GBIF Secretariat*. <https://doi.org/10.15468/39omei> (2021).
- Mesibov, R. Archived websites: A Data Cleaner's Cookbook (version 3) and all BASHing data blog posts 1–200. *Zenodo* <https://doi.org/10.5281/zenodo.6423347> (2022).
- Chamberlain, S. A. & Boettiger, C. R. Python, and Ruby clients for GBIF species occurrence data. *PeerJ Preprints* **5**, e3304v3301 (2017).
- Chamberlain, S. *et al.* *rgbif*: Interface to the Global Biodiversity Information Facility API. R package version 3.7.1. Available from <https://cran.rproject.org/package=rgbif> (2022).
- Nguyen, N. H. *et al.* FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).
- Sousa, D. *et al.* Tree canopies reflect mycorrhizal composition. *Geophys. Res. Lett.* **48**, e2021GL092764 (2021).

53. R Core Team. R: A language and environment for statistical computing. <https://www.R-project.org/> (2020).
54. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. <https://ggplot2.tidyverse.org> (2016).
55. Bederson, B. B., Shneiderman, B. & Wattenberg, M. Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies. *ACM Trans. Graph.* **21**, 833–854 (2002).
56. Simpson, H. J. & Schilling, J. S. Using aggregated field collection data and the novel r package fungarium to investigate fungal fire association. *Mycologia* **113**, 842–855 (2021).
57. Robertson, T. *et al.* The GBIF Integrated Publishing Toolkit: Facilitating the efficient publishing of biodiversity data on the Internet. *PLoS One* **9**, e102623 (2014).

Acknowledgements

The authors thank all members, volunteers, and guests of NJMA who participated in forays throughout the years. The list of most active participants is incomplete and we feel the deepest gratitude for everyone who was unintentionally left out: James Barg, Stefanie Bierman, Elizabeth Broderick, John Burghardt, Nina Burghardt, Keara Giannotti, Marc Grobman, Lynn Hugerich, Paul Hugerich, Kumiko Itagaki, Albert Karvois, Richard Kelly, Mark Michleski, Sue McClary, Maria Niciporciucas, Paul Patterson, Maricel Patino, James Richards, Jeremiah Smithson, Dorothy Smullen, Virginia Tomat, Betty Wise, Edward Zeme, with special thanks to Igor Safonov for the support throughout our project. We thank Evgeniy Meyke for his incredibly helpful assistance with EarthCape. We thank Sergey Bolshakov from Komarov Botanical Institute of the Russian Academy of Sciences for the help with data validation. We are grateful to Dr. Dmitry Schigel for critical reading of the manuscript. This work was supported by RTR2019–2022 grant from Kean University.

Author contributions

Data collection and digitizing: M.S., S.B., data analysis: N.L., N.F., L.L. and C.A., data deposition: N.F., project supervision: M.S. All authors contributed to the preparation of the manuscript.

Competing interests

Authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023