An assessment of biodiversity data shortfalls for ectomycorrhizal fungi in Canada

2026-09-22

AUTHORS:

Isaac Eckert[1,2] isaac.eckert@mail.mcgill.ca, ORCID 0000-0002-4384-5041

Justine Karst[3] karst@ualberta.ca, ORCID 0000-0003-0497-1552

Monika S. Fischer[4,5] monika.fischer@botany.ubc.ca, ORCID 0000-0001-5671-5572

Stephanie N. Kivlin[6] skivlin@utk.edu, ORCID 0000-0003-2442-7773

Clara Qin[7] claralqin@gmail.com, ORCID 0000-0002-9561-4704

Brontë R. Shelton[8,9] bronte.shelton@ubc.ca, ORCID 0000-0002-6915-2307

Diego A. Yusta Belsham[8,9] dayusta@gmail.com, ORCID 0009-0009-4930-8608

Jason Pither[8,9,*] jason.pither@ubc.ca, ORCID 0000-0002-7490-6839

Affiliations:

[1] Department of Biology, McGill University, Montréal, Québec, Canada

[2] Quebec Centre for Biodiversity Science, Montréal, Québec, Canada

[3] Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada

[4] Department of Botany, Faculty of Science, University of British Columbia, Vancouver, British Columbia, Canada

[5] Department of Forest and Conservation Sciences, Faculty of Forestry and Environmental Stewardship, University of British Columbia, Vancouver, British Columbia, Canada

[6] Ecology and Evolutionary Biology Department, University of Tennessee, Knoxville, Knoxville, Tennessee, USA

[7] Society for the Protection of Underground Networks (SPUN), Dover, Delaware, USA

[8] Department of Biology, Irving K. Barber Faculty of Science, University of British Columbia, Kelowna, British Columbia, Canada

[9] Okanagan Institute for Biodiversity, Resilience, and Ecosystem Services, Irving K. Barber Faculty of Science, University of British Columbia, Kelowna, British Columbia, Canada

* Corresponding author: Jason Pither (jason.pither@ubc.ca).

Abstract

Ectomycorrhizal (EcM) fungi play critical roles in global carbon and nutrient cycling. In Canada, vast forest ecosystems dominated by EcM trees are experiencing rapid environmental and anthropogenic change that threaten the symbiosis. Here we quantify and discuss seven biodiversity data shortfalls that constrain our knowledge of EcM fungi in Canada, and impede the development of effective conservation plans. We based our assessments on publicly accessible sequence data from GlobalFungi and GenBank, filtered to samples within Canada. We found **2,822** unique EcM fungal species hypotheses, of which only **58**% could be assigned a species name. The **1,411** unique named species comprised **124** genera, roughly **38**% of the genera known globally. An overriding shortfall is that samples are lacking from large areas of the country, and with respect to both abiotic and biotic niche space. For example,

root-derived host information is missing or unresolved for **79**% of the named species, and among **1,002** species georeferenced in GlobalFungi, 50% have **5** or fewer occurrence records nationally. Although our assessment highlights large biodiversity data shortfalls for EcM fungi in Canada, it also provides the foundation for a coordinated, pan-Canadian sampling effort that can be optimized to address multiple shortfalls simultaneously.

Keywords

ectomycorrhizal fungi, species hypotheses, biodiversity shortfalls, fungal conservation, DNA metabarcoding, boreal forest

Introduction

Large scale biodiversity data are often incomplete and unrepresentative, owing to several shortfalls that limit ecological and evolutionary knowledge and affect land stewardship (conservation, management, restoration) (Hortal et al. 2015). These shortfalls include gaps in knowledge about formal species descriptions (Linnean), species' geographic distributions (Wallacean), abundance and population dynamics (Prestonian), evolution of species lineages (Darwinian), species traits and functions (Raunkiæran), abiotic tolerance limits (Hutchinsonian), and species interactions (Eltonian).

Significant progress has been made in addressing these shortfalls for animals and plants, yielding extensive new data to inform conservation and management strategies. For example, billions of new species observations made over the past decade have enabled efforts to estimate

the geographic ranges of nearly all vertebrates, many plant species, and even an increasing number of arthropods (www.inaturalist.org; Lumbierres et al. 2022; Daru 2024; Daru 2025; Moulatlet et al. 2025). The IUCN Red List, which curates a variety of biodiversity data to perform extinction risk assessments, has now assessed over 90% of extant vertebrates and provides a wealth of standardized information on species geographic distributions, habitat preferences, elevational ranges, threats, and demographic trends (IUCN 2025). In addition, global efforts to collate data on species functional traits, genomes, and biotic interactions have produced or expanded large open access databases (Poelen et al. 2014; Kattge et al. 2020; Schmidt et al. 2020; Karachaliou et al. 2025; Matias Rodrigues et al. 2026), enabling researchers to predict vertebrate food webs (Hao et al. 2025) and trait-environment relationships (Li and Prentice 2024) at global scales, resolve megaphylogenies (Smith and Brown 2018; Upham et al. 2019; Kawahara et al. 2023), and track widespread genetic response to ongoing global change (Shaw et al. 2025). Even some of the most challenging shortfalls such as the Prestonian (i.e., population abundance and dynamics) are being filled. Updated global abundance tracking datasets like BioTIME (Dornelas et al. 2025), PREDICTS (Hudson et al. 2017), and the Living Planet Index (Ledger et al. 2023) are routinely relied on to detect and attribute biodiversity change, and taxa-specific efforts have identified widespread declines in what are historically data deficient groups like butterflies (Edwards et al. 2025) and other insects (Van Klink et al. 2024). These efforts to address major biodiversity data shortfalls, aside from increasing our biodiversity knowledge writ large, have directly benefited both applied and predictive research to inform the conservation of Earth's biodiversity amid accelerating global change (Steinke et al. 2025).

Far less progress has been made, however, in addressing these shortfalls for fungi (Mueller et al. 2022), including ectomycorrhizal (EcM) fungi. EcM fungi form obligate associations with the majority of tree stems on earth (Steidinger et al. 2019) as well as many shrub and herbaceous species that are widespread in alpine and tundra ecosystems globally (e.g. *Salix*, *Betula*, *Bistorta*, *Dryas*; Timling et al. (2012)). EcM fungi mediate soil nutrient and carbon cycling (Fry et al. 2019; Hawkins et al. 2023), are associated with forest productivity (Anthony et al. 2022), and promote expansion of plant ranges (Nuñez et al. 2009; Nuñez and Dickie 2014; Policelli et al. 2023). EcM fungi also are sensitive to shifts in forest management (Tomao et al. 2020; Lindahl et al. 2026) and climate change (van Nuland et al. 2024). Despite their importance and sensitivity to the drivers of global change, fungi have only recently been considered in conservation assessments or plans (Bazzicalupo et al. 2022; Niskanen et al. 2023; van Galen et al. 2025; van Nuland et al. 2025; Qin et al. 2026). While biodiversity assessments have become increasingly common and our ability to accurately predict global distributions of fungi is improving (Tedersoo et al. 2010; Tedersoo et al. 2012; Tedersoo et al. 2014; Větrovský et al. 2019; Větrovský et al. 2020; Tedersoo et al. 2021; Tedersoo et al. 2022; Qin et al. 2023; Abrego et al. 2024; Delhaye et al. 2024; Ovaskainen et al. 2024; van Galen et al. 2025; van Nuland et al. 2025; Qin et al. 2026), there remains a considerable gap in ecologically and evolutionarily relevant data required to inform broadscale conservation (Cameron et al. 2018; García-Roselló et al. 2023; Li et al. 2023; Mikryukov et al. 2023; Qin et al. 2023; Copot et al. 2024).

With the aim of inspiring and directing progress to gather the biodiversity data required to inform conservation and management, here we assess each of seven key shortfalls for

EcM fungi in Canada. Canada hosts around 12% of the Earth's forests (spanning 40% of Canadian land, an area equivalent to six times the size of France; (Natural Resources Canada 2022)), 24% of Earth's remaining intact ecosystems (Allan et al. 2017), and 20% of global soil carbon stocks in the top 1m (Sothe et al. 2022). With the majority of trees in Canada requiring symbiosis with EcM fungi (cf. Taylor et al. 2000; Steidinger et al. 2019), and large expanses of arctic tundra hosting EcM plants (Timling et al. 2012; Parisy et al. 2025), the conservation of EcM fungi across Canada carries planetary consequences for biodiversity and critical ecosystem services like carbon cycling.

For each of the seven shortfalls, we discuss past progress in gathering data on EcM fungi, and also consider data requirements that distinguish fungi from other groups like animals and plants, when appropriate. Finally, we estimate the breadth of current data gaps, laying a foundation for future assessments to track progress on addressing the shortfalls.

Data and Methods

Detailed methods are provided in the Supplemental Materials SM1 (Section **1**). Briefly, our assessments draw on publicly accessible fungal sequence records from GlobalFungi v5 [https://globalfungi.com/; Větrovský et al. (2020)] and NCBI GenBank (accessed 29 June 2026), via targeting internal transcribed spacer (ITS) sequence of the small-subunit ribosomal RNA. GlobalFungi is a curated database of community-survey amplicon data, consistently processed but aggregated from studies with varied protocols. Soil and soil-derived substrates account for **56.3**% of its **84,972** samples and roots for **18.7**%; in Canada, however, roots make up only **4.6**% of samples, and **3.0**% of those carrying EcM taxa. GenBank is a

general-purpose repository spanning diverse studies and sampling efforts, including targeted root sampling. For GlobalFungi we retained ITS2 and ITS1+ITS2 amplicon records and used the pre-assigned fungal species hypothesis (SH) codes and taxonomy that GlobalFungi provides from UNITE v10 [https://dx.doi.org/10.15156/BIO/2959332; 4 April 2024; Nilsson et al. (2019)]. In their pipeline, representative sequences meet a 98.5% similarity threshold for each SH. As noted by van Galen et al. (2025), GlobalFungi includes samples from the Global Soil Mycobiome Consortium (Tedersoo et al. 2021), 12 of which are located in Canada. For processing GenBank records, we aimed to match the GlobalFungi pipeline as closely as possible. We retrieved all fungal ITS records for Canada and used ITSx (Bengtsson-Palme et al. 2013) to detect and extract the ITS2 sub-region, discarding records with no detectable ITS2 or with an ITS2 fragment shorter than 100 bp. This retains only records carrying ITS2, matching the ITS2/ITSboth barcoding_region restriction applied to GlobalFungi. We assigned SH codes and taxonomy to the extracted ITS2 fragments using the same UNITE release (above) and vsearch v2.30.4 (Rognes et al. 2016) at 98.5% global-alignment identity. Because GenBank's content and assigned taxonomy change continuously, all GenBank-derived counts reported here reflect the state of the database on the access date; the retrieved record set is archived with the code and data (YYY) to ensure reproducibility.

We combined the GlobalFungi and GenBank-based SH and taxonomy datasets, then assigned the sequences to EcM guilds based on genus-level guild assignments in the FungalTraits database (Põlme et al. 2020). For any analyses that required georeferenced samples, we binned sample coordinates to three decimal places of latitude and longitude (~100 m at temperate latitudes), and these are henceforth referred to as sample "sites".

Methodological details for each shortfall analysis are provided in the Supplemental Materials SM1 (Section **1.4**). Code for our entire workflow is publicly available (XXX) and aims to be computationally reproducible. All input and output data are also archived and publicly available at YYY.

Table 1 summarizes the key properties of our Canada-specific GlobalFungi dataset, alongside comparator figures derived from all GlobalFungi samples worldwide. Table 2 summarizes relevant information about the GenBank data. Figure 1 shows the locations of both GlobalFungi and GenBank samples in Canada (**486** GlobalFungi and **1,180** GenBank unique locations).

Linnean shortfall

The Linnean shortfall is operationalized as the discrepancy between the number of species that have been formally described and the number of species that exist or have ever existed (Lomolino 2004). The magnitude of this shortfall is difficult to quantify for EcM fungi, since it depends on the total number of species (which is generally unknown) as well as our ability to delineate and describe new species (Hortal et al. 2015). While EcM fungi have been traditionally delineated based on morphological characteristics using approaches like microscopy, the advent of genetic sequencing and high-throughput technologies completely transformed fungal systematics (Miralles et al. 2020) and greatly increased species description rates (Cunningham et al. 2024). These molecular techniques and the corresponding species concepts (referred to as species hypotheses) have been widely adopted; in 2018 over 94% of studies describing new fungal taxa relied on molecular techniques (Miralles et al. 2020). Cunningham et al. (2024) predict that Kingdom Fungi could see an additional 64,000–74,000

newly described species by the end of the century, thanks to high-throughput workflows. A recent, conservative estimate of the global diversity of EcM fungi is 25,499 (+/- 54) – 55,500 (+/- 518) species based on sequence datasets from GlobalFungi and the Global Soil Mycobiome consortium, respectively (van Galen et al. 2025). However, these estimates represent lower-bound estimates and are highly uncertain for reasons discussed below.

Despite this progress, the dependency on molecular species concepts means that almost all newly identified species remain undescribed, or "dark", taxa. Many identified species have only been detected a handful of times, often solely by molecular barcoding, and often only by single sequences (Cazabonne et al. 2024). In both GlobalFungi (Větrovský et al. 2020) and the Global Soil Mycobiome (Tedersoo et al. 2021) databases, a minority of taxa ($< 25\%$) are assigned species-level names. As a result, when it comes to assessing existing knowledge for fungi, global assessments like the IUCN Red List have evaluated only a tiny fraction of described species (Mueller et al. (2022) estimated 0.4%, a percent that is now larger but still much less than 5%). The requirement that species be described before qualifying for Red List assessment represents a serious bottleneck, and contributes to why fungi are the most poorly assessed group across all Eukaryotes excluding protists (Niskanen et al. 2023; IUCN 2025).

Quantifying and addressing the Linnean shortfall for EcM fungi has also been hindered by the lack of standardized and integrated global taxonomic repositories. Unlike for vertebrates and plants, comprehensive and harmonized taxonomies for EcM fungi rely on databases that combine more traditional species concepts (e.g., morphological) with modern approaches (e.g., molecular). Fortunately, progress has been made on this front. Originally released in 2003, the publicly accessible UNITE database, which offers standardized fungal taxonomies

based on molecular records, has grown rapidly over the past decade (Abarenkov et al. 2024). For example, since 2014, the number of fungal species (species hypotheses) represented in UNITE has more than tripled. Importantly, as of 2019, UNITE has started linking molecular sequences (e.g., metabarcodes) to standardized species hypotheses harmonized against the Global Biodiversity Information Facility (GBIF) backbone taxonomy, thereby facilitating their use in informing policy. These advances have been instrumental in standardizing species concepts across EcM fungal research, increasing reproducibility and reusability of data, and facilitating crosstalk with other global databases that leverage UNITE.

Our assessment of the Linnean shortfall for EcM fungi in Canada starts with a basic accounting of EcM fungal taxonomic data. We emphasize at the outset that the sampling footprint in Canada is geographically biased and clustered (Figure 1), especially with respect to the potential abiotic and biotic niche space of EcM fungi (see Hutchinsonian and Eltonian shortfalls sections), so all results should be interpreted with this in mind. Within the combined GlobalFungi and GenBank dataset there were **2,822** unique UNITE SH codes, and the **2,062** codes from the Canada-specific GlobalFungi dataset represents **16**% of the **12,970** EcM fungal SH codes within the entire GlobalFungi dataset.

Among the **2,822** unique EcM fungal SH codes, **1,629** (**58**%) were assigned to a UNITE named species, and **1,411** of these were unique (some SH codes are assigned to the same species epithets). Of the named SH codes, **687** were shared between GlobalFungi and GenBank, **436** were unique to GlobalFungi, and **506** were unique to GenBank.

The **1,411** named species comprise **124** unique genera, which is around **38**% of the estimated **327** EcM fungal genera known globally.

The difference between the **2,822** unique SH codes and the **1,629** with names attached represents the complement of "dark taxa" specific to our dataset: about **42**% (**1,193**) of codes in the combined GlobalFungi + GenBank dataset have no match above the UNITE species-level assignment threshold, and thus cannot be assigned a species name at this time. This figure is not directly comparable to the OTU-level enumerations of dark taxa recently reported by van Galen et al. (2025), as they used a custom bioinformatics pipeline that, for example, used a 97% similarity threshold for taxonomic assignment rather than the 98.5% used by GlobalFungi (and by us). Nevertheless, for reference, we (i) provide a Canada-specific map of the estimated percentage of EcM fungal OTUs within local assemblages that are dark taxa, adapted from the geospatial model of van Galen et al. (2025) (Supplemental Materials Figure S1), and (ii) accessed their sample-specific estimates of the proportion of OTUs unassigned to species (available at: https://doi.org/10.6084/m9.figshare.28830371), which is analogous to our estimates described above, and calculated an average of **57**% among their **1,368** Canadian forest and tundra samples (the most comparable subset to ours).

The Linnean shortfall for EcM fungi in Canada is clearly substantial, but also challenging to estimate accurately because the available sample data are (i) so few relative to the target land area, (ii) geographically clustered, and (iii) restricted to a small portion of the potential EcM fungal niche space (Figure 1; see also Hutchinsonian, Eltonian, and Wallacean shortfalls). The same limitations apply to the full GlobalFungi dataset (see global sampling density map, Figure S2), upon which the estimate of 25,500–55,500 EcM fungal species globally was based (van Galen et al. 2025).

Moreover, and the reason we do not provide estimates of sample coverage ($\hat{C}$) or EcM fungal

taxon richness in Canada, the available estimation methods make assumptions that cannot be met by GlobalFungi or comparable community eDNA data. Specifically, Chao1, Chao2, and $\hat{\mathrm{C}}$ (Chao 1987; Chiu and Chao 2016) all derive from the same underlying quantities: the number of rare taxa observed exactly once (f_1) or twice (f_2) in a sample (Chao1), or across sites (Q_1 and Q_2, the incidence-based analogue, for Chao2). Coverage is, in its simplest form $\hat{C} = 1 - f_1/n$, where "n" represents the total sample reads. All of these methods assume that f_1 and f_2 (or Q_1 and Q_2) faithfully reflect how much rare diversity remains undetected in the true underlying community. This assumption cannot be met with GlobalFungi data because, when estimating richness of specific taxonomic guilds (like EcM fungi), the guild-specific read count of every GlobalFungi sample is the outcome of an upstream taxonomic (EcM fungal genus) filtering step applied to each sample's whole-community sequencing output. To illustrate, across the **1,388** soil-only Canadian GlobalFungi samples with at least one EcM fungal read, a median of **94**% (IQR **79**–**99**%) of total sample sequencing depth is discarded by the EcM filtering step before any richness estimator would see the data (**Figure S3**). It is therefore problematic to assume that the f_1 and f_2 (or Q_1 and Q_2) values that come from the remaining EcM fungal reads represent what the methods assume they represent. Notably, this problem is separate from the widely-recognized challenges that PCR / sequencing methodologies pose for diversity estimation (Boshuizen and Te Beest 2023; van Galen et al. 2025), and from other problems highlighted in a recent study (Elena Schmitz and Rahmann 2025).

Wallacean shortfall

The Wallacean shortfall encompasses our lack of knowledge of where species occur (i.e., species geographic distributions). For fungi, the Wallacean shortfall was once considered negligible, as longstanding hypotheses predicted that microbial species were ubiquitous across space (reviewed in Green et al. 2008). The advent of molecular methods enabling whole-community sequencing of fungal diversity, however, has shown the opposite. Field surveys detecting putatively endemic fungi are common (Talbot et al. 2014; Kivlin and Hawkes 2020; Bazzicalupo et al. 2022), including in Canada or western North America (Kranabetter et al. 2018), and studies have demonstrated geographically constrained dispersal for most EcM fungal spores (Peay et al. 2012; Abrego et al. 2024). Thus, although some EcM fungal taxa may be distributed over large areas (e.g. circumpolar; Geml et al. (2015); Noffsinger et al. (2026)), many more taxa likely exhibit more constrained distributions, mirroring what is observed among other taxonomic groups (Brown et al. 1996).

In recent years, species distribution models (SDMs), which relate species occurrences to abiotic and sometimes also biotic predictors to estimate species geographic distributions, have increasingly been used to predict the potential distributions of EcM fungi (Qin et al. 2023; van Nuland et al. 2025; Qin et al. 2026). For example, in the most comprehensive modeling of EcM fungal species distributions thus far, Qin et al. (2026) used Hierarchical Modelling of Species Communities (HMSC) (Tikhonov et al. 2020), informed by worldwide GlobalFungi soil samples binned into 30 arc-second grid cells (~1 km × 1 km), to model the distributions of 2,669 EcM fungal species hypotheses (based on a tailored pipeline that used OTUs composed of sequences clustered to 97% similarity). This equates to approximately

10% of the estimated minimum number of EcM fungal species globally (see above). Qin et al. used a minimum occurrence cutoff of 31 grid cells, which is consistent with common recommendations (e.g. van Proosdij et al. 2016).

For comparison, and serving as our first, straightforward measures of the Wallacean shortfall, we highlight that among the **2,062** unique SH codes that were georeferenced in Canada (from GlobalFungi samples exclusively), **969** (**47**%) have fewer than 30 occurrence records globally, while **1,855** (**90**%) have fewer than 30 occurrence records within Canada (Table 3).

Our second approach to quantifying the Wallacean shortfall focuses on the frequencies of occurrence (or detections) of EcM fungal taxa more generally. Figure 2 shows histograms of those frequencies, enumerated within Canada (panels a and b) and globally (c, d) for both named species (a,c) and SH codes (b, d), again limiting analyses to the GlobalFungi dataset and sites equating with 30 arc-second grid cells. The histograms show that most taxa have been detected in few sites whereas a few have been detected in many sites. However, unlike for macro-organisms like mammals and birds, sampling is currently insufficient to estimate the true frequency distributions of occurrence, at any spatial scale, for EcM fungi native to Canada, even among taxa that are putatively endemic to Canada or western North America (Kranabetter et al. 2018). Together with Figure 2, the numbers presented in Table 3 provide a straightforward picture of the data limitations that underlie the substantial Wallacean shortfall for EcM fungal taxa in Canada.

Prestonian shortfall

Of the seven shortfalls, the Prestonian, which represents our lack of knowledge about the abundance of species, is often most pronounced (Hochkirch et al. 2021). This is especially true for fungi, since abundance trends and population dynamics over space and time remain mostly unknown (Koch et al. 2018; Martinović et al. 2021). Evidence from vertebrate communities suggests that most populations require relatively continuous monitoring over decadal timespans to provide robust estimates of abundance change (White 2019). In contrast, fungi often exhibit fast generation times and rapid community turnover (Ning et al. 2021); thus, resolving population dynamics likely requires fungal-specific sampling approaches (Tóth and Barta 2010). Short-term population dynamics (resolved via sequencing host root tissue) can incorporate stable-isotope probing via isotopically-labeled carbon substrates or water to track actively growing EcM fungal taxa (Koch et al. 2018) on the order of days to weeks. Fungal turnover in community composition during intermediate time scales can be addressed via repeated sampling (Kivlin and Hawkes 2020; Martinović et al. 2021) and may be important for understanding succession or response to disturbances. Long-term population dynamics meant to indicate factors like extinction risk are likely more complicated and could involve the monitoring of spore deposition (Kranabetter et al. 2026) or spore banks which can sustain viable propagules for many years (Nguyen et al. 2012; Pither and Pickles 2017), identifying genetic structures of fungal populations though whole genome sequencing (Vincenot and Selosse 2017), and/or incorporating phylogenetic methods to detect radiations versus extirpations (Beaulieu and O'Meara 2016). Unfortunately, studies with repeated sampling designs over both short and long time scales are uncommon for EcM fungi.

Correspondingly, estimates of fungal persistence within a single location, much less across a species range, are exceedingly rare. Even when samples are collected repeatedly over time (Averill et al. 2019), the destructive nature of chemical assays, microscopic assays, and DNA sequencing of soils and/or roots means that temporal estimates of fungal persistence must account for a spatial component as well (Carini et al. 2020).

Unlike other taxa, approaches taken to measure the abundance of EcM fungi vary considerably, which makes aggregating data across regions, species, and studies challenging. For example, fungal abundance can be represented as estimates of total biomass detected from chemical signatures (e.g., phospholipid fatty acids in (Olsson 1999); ergosterol in (Ekblad et al. 1998); biomass inferred from carbon use efficiencies in (Jörgensen et al. 2025)) or microscopic counts of EcM fungal root tip colonization (Soudzilovskaia et al. 2015). Minirhizotron tube measurements that track mycorrhizal fungal growth over time are rare but those estimates suggest that EcM fungal hyphae turn over on the order of weeks to months (Treseder et al. 2010; Allen and Kitajima 2013; Defrenne et al. 2021). Estimates of fungal abundance based on DNA sequences (e.g., qPCR and next generation sequencing) are often biased by unequal target gene copy number (Taylor et al. 2016), PCR bias against certain EcM fungal clades (Janowski and Leski 2023), and the fact that the DNA concentration of fungal biomass varies across growthforms (e.g., aseptate arbuscular mycorrhizal fungi). Relative abundance estimates of EcM fungi compared to total read counts suffer from the fact that the former are a simplex (sum to 1) and often are not corrected for total unequal sequencing depth among samples (see above), sequencing technologies, and/or gene regions (Weiss et al. 2017; Quinn et al. 2019). In the future, estimates of fungal abundance with deeper next

generation sequencing may more closely approximate biomass, and targeting sequencing of whole genomes may allow us to estimate the number of individuals in a sample more reliably.

Because it is difficult to estimate the number of fungal individuals in a sample, it is nearly impossible to model population dynamics of most EcM fungal lineages. Demographic models require tracking individuals through survival, growth, and reproduction over time. But, even when sampling of EcM fungi is repeated at monthly time intervals, the same individuals may not be detected. For the purpose of conservation monitoring, Dahlberg and Mueller (2011) propose fruiting body counts as an operational definition for population size; however, to be reliable, they suggest that EcM fungal population trends should be assessed over at least 50 years. Instead, a focus on genetic diversity over time and among populations across an EcM fungal species' range may be the most promising avenue forward to understand persistence of fungal populations (Douhan and Rizzo 2005; Abe et al. 2024). Using this approach, Golan and colleagues demonstrated that genets of the EcM fungus *Amanita phalloides* can persist for more than three decades across multiple locations (Golan et al. 2024). As longitudinal data accumulate for EcM fungi, future studies can begin to incorporate changes in abundance into population demographic models, at least as stage-based estimates of population growth and longevity.

To quantify the Prestonian shortfall for EcM fungi in Canada, we used the BioTIME database, which represents the largest and most comprehensive effort to compile abundance data (Dornelas et al. 2025). Yet, despite decent coverage for other kingdoms, BioTIME includes only a single study that tracks fungal abundance over time based on data collected from Bavarian Forest National Park in south-eastern Germany (Thorn et al. 2016) over a

period of 5 years. Of the **110** fungal species included in that study, two have been recorded in Canada. Here, we operationalize the Prestonian shortfall as the percentage of the **1,411** named EcM fungal species present in Canada that lack some form of abundance time series data recorded anywhere on Earth. We estimate this shortfall to be **1,409** out of **1,411**, or **99.9**%.

Darwinian shortfall

The Darwinian shortfall describes the lack of knowledge about how species and their traits have evolved (Diniz-Filho et al. 2013). Traditional species concepts have been especially problematic in microbes as criteria for delineating species differ among groups of fungi (Cai et al. 2011; Kwon-Chung et al. 2017; Xu 2020; Janowski and Leski 2023); and previously agreed-upon phylogenies can rapidly and drastically shift (Tedersoo et al. 2010; Tedersoo et al. 2014; Niskanen et al. 2023).

The hyperdiversity and short generation times of microorganisms (including fungi) can make resolving phylogenetic relationships more challenging than in other organisms that lack abundant horizontal gene transfer, multiple genomes, high phenotypic plasticity, or increased potential rates of mutation accumulation (James et al. 2020; Xu 2020; Li et al. 2021; Janowski and Leski 2023). Although the discussion of the evolution of the EcM fungal habit has been ongoing since at least the 1980s (Malloch et al. 1980; Read 1991), the topic has continuously been revisited and revised (Hibbett et al. 2000; Matheny et al. 2009; Sato 2024). The timing and geographic origins of the original emergence of EcM fungi are another area that has been limited by the lack of available taxonomic data and has been much discussed

and summarized (Ryberg and Matheny 2011; Looney et al. 2016; Niskanen et al. 2023; Sato 2024). Despite this ongoing research, there is still a lack of agreement on how many times the EcM symbiosis has independently evolved - from as few as 27 (Sánchez-García et al. 2020; Cai et al. 2024; Sato 2024) to upwards of 80 times (Tedersoo et al. 2010; Tedersoo and Smith 2013; Sánchez-García et al. 2014; Tedersoo et al. 2014; Tedersoo and Smith 2017; Strullu-Derrien et al. 2018) - and which fungal taxa truly belong in the EcM lineage (Tedersoo and Smith 2013; Tedersoo and Smith 2017), and even which genes should be taken into account when building phylogenies (Sato et al. 2017; Xu 2020; Li et al. 2021; Cai et al. 2024).

The limited taxonomic data on EcM fungi necessitates a large amount of extrapolation and reliance on closely related taxa to infer relationships (Tedersoo et al. 2010; Tedersoo and Smith 2013). Many attempts to delineate phylogenies have been focused on single lineages (Lebreton et al. 2022; Kobayashi et al. 2023; Cai et al. 2024; Noffsinger et al. 2026), and there have been limited attempts to synthesize across studies (Tedersoo et al. 2010; Tedersoo and Smith 2013; Tedersoo and Smith 2017), impeding a broader understanding of the overall evolution. Additionally, the majority of attempts have been biased towards the Agaricomycetes class (Sánchez-García et al. 2014; Tedersoo et al. 2014; Cai et al. 2024; Sato 2024).

Given their symbiotic lifestyle, EcM fungal diversification could be linked to host diversification and host switching (Sato et al. 2017; Wilson et al. 2017; Strullu-Derrien et al. 2018; van der Linde et al. 2018; Sato and Toju 2019; Cai et al. 2024; Sato 2024). Time since major disturbance, such as the Cretaceous–Paleogene (K–Pg) boundary or global cooling events

in the Late Paleogene–Neogene (Strullu-Derrien et al. 2018), and functional traits, such as fruiting body forms (Varga et al. 2019; Sánchez-García et al. 2020) or the formation of spores (Sato 2024), are also potential drivers of EcM fungal diversification. Due to a lack of available taxonomic data and the difficulties in resolving fungal phylogenies, function or trait-based approaches might be a substitute for evolutionary relationships (Aguilar-Trigueros et al. 2015; Zanne et al. 2020). However, context-dependence and the lack of agreed upon life-history strategies and functional groups for EcM fungi, combined with the fact that there is not a standardized set of morphological or functional data collected in EcM studies (see Raunkiaeran shortfall section), has made this approach challenging. For an approach to be viable, studies must move beyond metabarcoding to investigate the functional roles of EcM fungi, and a standardized framework for studying and recording trait data must be implemented.

More recently, there has also been growing interest in the rate of modern EcM fungal evolutionary processes (Dauphin and Peter 2024). While there is evidence of fungal evolution on ecologically-relevant timescales, such as adaptation to heavy metals (Bazzicalupo et al. 2020) as well as temperature and precipitation (Sønstebø et al. 2022), these traits are still relatively poorly characterized, inhibiting effective conservation strategies (Dauphin and Peter 2023; Dauphin and Peter 2024). This is another area that will require studies to move away from metabarcoding and towards whole genome sequencing to better investigate speciation, as well as changes in functional genes that barcoding cannot capture (Shelton et al. 2026). For example, *Suillus* species are emerging as model organisms due to the availability of a high quality genome and detailed knowledge of life cycles (Branco et al. 2015; Bazzicalupo et

al. 2020; Dauphin and Peter 2024; Lofgren et al. 2024). Overall, fungi lag behind in whole genome studies and, while there has been an increasing number of whole fungal genome sequencing initiatives, the number for EcM fungi in particular is still low with an estimated 0.5% of the 50,000 known EcM fungi genomes sequenced (Dauphin and Peter 2024).

An increase in whole genome sequencing, as well as improved trait characterization, are necessary to move beyond a single model organism and fully capture the large genetic and functional diversity of EcM fungi. However, whole genome sequencing may be prohibitively expensive and computationally intensive for some studies. In this case, sequencing for conserved functional genes of interest, such as those relating to melanin production or heavy metal tolerance, in addition to traditional ITS sequencing could help to elucidate the potential niche and functionality of EcM fungal communities at a lower cost (Suenaga 2015; Dauphin and Peter 2023). Strategies in landscape genomics that have been developed for other organisms can also be well-applied to EcM fungi, as noted in Dauphin and Peter (2024), such as targeting gaps in spatial and temporal resolution and employing stratified sampling to increase statistical power (Dauphin and Peter 2023).

Quantifying the Darwinian shortfall is challenging, since approaches to gather and analyze evolutionary data change over time. Here, we decided to focus on the representation of Canadian EcM fungal species in terms of available genomic data. While many fungi, including EcM, are often delineated based on reference barcodes of commonly sequenced regions like ITS (nuclear ribosomal) and CO1 (mitochondrial), these highly conserved regions of the genome are poorly suited to understanding the evolutionary relationships among species (Dentinger et al. 2011), which is best done using whole genome or long-read approaches

(Varga et al. 2019). For fungi, the MycoCosm genomic database (Grigoriev et al. 2014) is the primary source of compiled genomic data. Using this database, we quantified the Darwinian shortfall as the percentage of the **1,411** named EcM fungal species in Canada absent from the database. This amounted to **1,344** species, which represents **95.3**%, indicating this shortfall remains severe.

Raunkiæran shortfall

The Raunkiæran shortfall is the lack of data on species traits and functions. Characterizing species by functional traits instead of phylogenetic relatedness is important for understanding how species contribute to ecosystem functions and interactions (Ackerly and Cornwell 2007; Aguilar-Trigueros et al. 2015; Kardol et al. 2016; Fry et al. 2019). Importantly, and more so in fungi compared to other kingdoms, phylogenetic relatedness does not always correlate well with the ecological or functional roles that species perform (Martiny et al. 2013). For example, while fungi in the genus *Cenococcum* (Ascomycota) have many functional, ecological, and genomic traits in common with EcM fungi in phylum Basidiomycota (Peter et al. 2016), fungal species within the same genus can exhibit considerable functional diversity (Carrillo et al. 2025). This disconnect between phylogeny and function highlights the need to collect functional trait data to connect fungal species and sequences with ecosystem functioning to better understand the vital role of EcM fungi in ecosystems across Canada.

The large diversity of functions across fungi complicates efforts to standardize trait data collection. Even among mycorrhizal fungal researchers, there are various approaches to collecting and reporting trait data. While researchers of arbuscular mycorrhizal (AM) fungi

have adopted Grime's C-S-R framework (Chagnon et al. 2013) to organize and standardize measures of functional traits like life history strategy across multiple dimensions (hosts, soil, fungal ecology; Antunes et al. (2025)), no such framework exists for EcM fungi. Assigning exploration types to EcM fungi holds promise, especially as these traits appear to covary with capacities for decomposition and nutrient cycling (Mansfield et al. 2026). However, traits focused on specific genes, enzymes, and/or resource flows can be context-dependent for both the environmental conditions and the specific mycorrhizal fungus being investigated. Thus, while unified trait databases are a valuable tool for summarizing broad patterns and generating testable hypotheses, producing these for EcM fungi remains a challenge.

Currently, FungalTraits is the most comprehensive functional traits database for fungi, with trait descriptions for over 160,000 fungal genera (Põlme et al. 2020), including 327 EcM fungal genera. Decades of field and laboratory observations form the basis of these trait descriptions that broadly describe fungal genera in terms of morphology (e.g., growth form, and/or mycelial exploration type) and ecology/interactions, which collectively point toward broad ecological roles for fungi.

We quantified the Raunkiæran shortfall as the coverage of trait information for EcM fungi in Canada, using six coarse ecological/morphological traits recorded in the FungalTraits database that are relevant to EcM fungi. A seventh trait, "growth form", is exclusively "filamentous mycelium" for all 327 EcM fungal genera, so is not considered further. It should be noted that the FungalTraits database does not include traits on enzymatic function or carbon acquisition, features critical to EcM fungal function.

As shown in Table 4, trait knowledge is concentrated in morphological characters derivable

from taxonomy (almost 100% coverage) and is near-absent for ecological-interaction traits. We acknowledge that the host-specificity trait is relevant to the Eltonian shortfall (below), but we discuss it here due to its inclusion in the FungalTraits dataset. The 6 genera with a documented host-specificity entry (all Pinaceae) are *Chroogomphus*, *Gomphidius*, *Pseudoboletus*, *Fuscoboletinus*, *Rhizopogon*, and *Suillus*. However, the first two of these genera appear to be primarily epiparasites of the latter two genera (Olsson et al. 2000), and based on morphology and multigene phylogenies, *Fuscoboletinus* has recently been reclassified as a subgenus of *Suillus* (Shi et al. 2025). *Pseudoboletus* is a monospecific genus whose single species (*P. parasiticus*) is a parasite of the earthball *Scleroderma citrinum* (though it has been shown to form ectomycorrhiza with *Pinus resinosa* roots in culture; Richter and Bruhn (1989)). The two ecological/interaction traits with near-zero coverage are documented for the same two genera, both members of the Endogonales (Mucoromycota): *Densospora* and *Jimgerdemannia* (Chang et al. 2019). *Densospora* is the only genus assigned a secondary lifestyle of "soil saprotroph" and an endophytic interaction capability of "root endophyte", while *Jimgerdemannia* carries the value "root-associated" for both fields. No other Canadian EcM genus has an entry for either trait, so these two characters provide essentially no resolving information across the assemblage. The Raunkiæran shortfall for Canadian EcM fungi is therefore primarily a shortfall in ecological-interaction knowledge specifically.

Hutchinsonian shortfall

The Hutchinsonian shortfall refers to the lack of knowledge about species tolerances to abiotic environmental conditions (Hortal et al. 2015). Addressing the Hutchinsonian shortfall will

not only improve our understanding of the factors underlying spatio-temporal variation in the abundance and distributions of EcM fungal species, but is also particularly relevant for predicting fungal responses to climate change (Steidinger et al. 2020).

For a taxonomic group as diverse as the EcM fungi, a species-level focus has not been the norm in primary research; instead, most studies assess community-level features such as species richness, composition, community aggregated traits, or total biomass in relation to environmental conditions. For example, soil properties can influence diversity and composition at fine spatial scales (Taylor et al. 2014), whereas climate typically emerges as the primary driver at broad spatial scales (Tedersoo et al. 2014; Větrovský et al. 2019; Mikryukov et al. 2023). A community-level focus has also been the classical approach in plant macroecology, which culminated in the creation of plant functional types and the development of dynamic global vegetation models in contrast to individual species response curves (Box 1996). While informative, these assemblage-level patterns provide limited insights into the abiotic niches of individual taxa. However, like the trajectory of plant biogeography, which has shifted from a community to species-level focus, EcM fungal biogeography has begun to shift research focus from describing community-level responses to characterizing individual species-level response curves via ecological niche models and SDMs (Davison et al. 2021; Kivlin et al. 2021; Qin et al. 2023; van Nuland et al. 2024; Qin et al. 2026). Given that the diversity and symbiotrophy of EcM fungi make it infeasible to estimate species' tolerances at scale via controlled physiological experiments, niche modeling will remain the primary means for estimating EcM fungal species' responses to abiotic conditions (but see Lofgren et al. (2024)).

To achieve best accuracy, niche models should be informed by sampling that captures all the

environmental conditions in which the target species occurs (Moudrý et al. 2024). This is especially important for presence-only occurrence data, for which niche modeling additionally requires that environmental conditions are sampled in proportion to their actual availability (Hastie and Fithian 2013). Thus, for our first assessment of the Hutchinsonian shortfall, we evaluate the degree to which available samples cover Canada's climate space (Figure 3a,d), defined here in simplified terms by mean annual temperature and mean annual precipitation. Figure 3 makes clear that this aspect of the Hutchinsonian shortfall is enormous for EcM fungi in Canada, particularly for the systematic samples from GlobalFungi (Figure 3b,e). For example, the most common climate conditions in Canada - those covering northern boreal, sub-arctic and arctic regions (Fig. 3a) - are among the most poorly represented among GlobalFungi (Figure 3e) and GenBank samples (Fig. 3f). GenBank samples appear biased towards areas with high mean annual temperature (Figure 3f), which are comparatively rare. The global sampling density map (Figure S2), which shows few samples across vast expanses of north-eastern Europe and northern Russia, implies similar climate sampling gaps at the global scale, with implications for EcM fungi SDMs based on global data (Qin et al. 2026).

The preceding assessment considered only climate. Evaluating sampling in relation to additional abiotic variables, such as soil and physiographic conditions, would undoubtedly reveal additional gaps, thus broadening the Hutchinsonian shortfall further. As a preliminary assessment of this, Figure S4 shows a map of GlobalFungi and GenBank samples overlaid on Canada's 15 ecozones, and Table S1 shows the sample numbers and sampling density (number of samples per 10,000 km^2) for each ecozone. Taken together with the climate space sampling results (Figure 3), these ecozone findings show that the Hutchinsonian shortfall is

extensive for EcM fungal taxa in Canada, and a priority will be designing sampling strategies that appropriately represent the potential abiotic niche space of species.

Eltonian shortfall

The Eltonian shortfall describes the lack of knowledge of biotic interactions. This is especially relevant for EcM fungi as they are primarily obligate symbionts, meaning that most species require a compatible plant host to reproduce. Here we first consider the Eltonian shortfall from the perspective of the EcM fungi, and specifically what data exist about potential hosts for the **1,411** named EcM fungal species in Canada. Given that the majority of Canada's forests and much of the tundra are dominated by EcM tree or shrub species, the arena within which functional EcM interactions could be realized is enormous. However, not all EcM plant species are suitable hosts for all EcM fungal species; EcM fungi exhibit a range of host breadths, from the comparatively narrow host-breadth of *Suillus*, which predominantly associates with members of the Pinaceae family, to the broad host-breadth of *Cenococcum geophilum*, which associates with a wide array of EcM tree species including gymnosperms and angiosperms (Molina and Horton 2015). While the mechanisms thought to influence host breadth are varied, and implicate processes at both the molecular and genomic level and landscape and regional scales (Molina and Horton 2015; Martin et al. 2016), much remains to be learned about why some EcM fungal taxa can colonize a wide range of hosts, while others are more restricted (Martin and Van Der Heijden 2024; Voller et al. 2024).

Demonstrating the existence of an EcM symbiosis is itself logistically demanding. The gold standard for evidence is demonstrating the presence of a Hartig net, and, once confirmed,

using DNA sampled from the ectomycorrhiza and host roots to identify the mycobiont and phytobiont partners (Brundrett and Tedersoo 2020). Such evidence is rare in practice. Most common are samples of environmental DNA of fungi (e.g. from soil samples) for whom the potential host species is inferred as the nearest plant, which can be problematic when there are multiple potential host species in the vicinity. Thus, for our first assessment of the Eltonian shortfall, we report that of the **1,469** GlobalFungi samples in Canada that included at least one EcM fungal SH code, **684** (**47**%) lacked information about potential hosts, i.e. the "dominant plant species" metadata field was empty. Of the **748** samples that did include potential host information and were collected from a plant-bearing substrate, **704** (**94.1**%) were collected from soil, and thus potential host information must be inferred. Only **44** samples were collected directly from roots (**5.9**%), and would therefore provide more confidence about the potential host identity, though these **44** samples yielded **848** fungal SH-code detections between them. A further **37** samples carried a dominant-plant label but were lichen samples, and are excluded from this denominator because the label does not indicate a root-associated host. Among the **9,403** GenBank sequence records, **7,582** (**80.6**%) were missing host information (in the "host taxon" field). GenBank records were treated differently from the outset, because the database has no controlled sample-type field: among the **1,821** records that did include host information, **1,075** (**59**%) were explicitly derived from roots or ectomycorrhizas, **144** (**8**%) demonstrably from soil, rhizosphere, or non-root tissue, and **602** (**33**%) gave no usable indication of the tissue sampled, **374** with an empty "isolation_src" field and **228** describing only the surrounding habitat (for example, "40 year-old Douglas-fir stand") rather than the material collected. Note that the **44** GlobalFungi figure counts *samples*, each of which yields many fungal detections, whereas the **1,075** GenBank figure

counts *records*, one sequence accession each; the two are not directly comparable.

Because host identity inferred from soil is unreliable, all fungus–host associations reported below rest on root-derived evidence only. For GlobalFungi this means root samples exclusively, so the **704** soil samples contribute to none of the association results. For GenBank it means only those records whose isolation-source field names root or ectomycorrhizal material; the remaining **746** records with host information are excluded from all association analyses, although they are retained in the taxonomic and geographic assessments (see Supplemental Materials SM1, Section **1.4.7**).

Next, we summarize the gap in what is known about potential EcM fungi – host interactions, again using information derived from the GlobalFungi and GenBank samples. On the fungal side, among the **1,411** named EcM fungal species, **1,116** (**79**%) lacked any associated host information. Among the **295** named EcM fungal species that were associated with at least one host species, the maximum number of unique host species associated with an EcM fungal species was **10**. If we expand the scope to include metadata from all global records from GlobalFungi and GenBank, then **767** (**54**%) of the **1,411** named EcM fungal species in Canada lacked any associated host information anywhere, and among the **644** species that were associated with at least one host species, the maximum number of unique host species associated with an EcM fungal species was **21**. Many samples list several co-occurring plants, so we retained every listed plant as a candidate host but distinguished associations supported by at least one record naming a single candidate plant (n = **214** host × fungal-species pairs, **19.4**%) from those resting entirely on records that named several (n = **891**). Restricting the analysis to the unambiguous evidence alone changes the conclusions only in degree: the

number of host species with any documented fungal partner falls from **20** to **18**, leaving **254** rather than **252** of the **272** native EcM host species undocumented.

On the host side, there were **59** unique host plant species documented in the metadata. For comparison, we cross-referenced data from the FungalRoot dataset (Soudzilovskaia et al. 2020, GBIF source: doi:10.15468/a7ujmj) with the BIEN database and its native species resolver (Maitner et al. 2018; Boyle et al. 2024; Enquist et al. 2026) to assemble a list of vascular plant species native to Canada that are EcM (see detailed methods in Supplemental Materials SM1, Section **1.3.3**). This yielded **272** species; **120** tree species, **119** shrub species, and **33** herbaceous species. Our estimate of **272** is conservative: species of *Acer*, *Fraxinus*, and *Juglans*, for example, are excluded from our list despite published evidence of genuine EcM associations for these genera elsewhere in the literature (Teste et al. 2020). FungalRoot's genus-level recommendation does not close this gap, since it independently classifies all three genera as AM. We therefore emphasize that our remaining Eltonian shortfall estimates, which rely on data about potential host species, are conservative estimates. Among the list of **59** unique host plant species names documented in the metadata, **28** matched names in our list of native EcM host plant species in Canada (Table S2). The remaining **31** names that did not overlap included ornamental species (which we did not consider), mis-spelled names, and other problematic entries (Table S3). Applying the root-evidence rule set out above, however, only **20** of those **28** matched names are supported by at least one root-derived record; these **20** are the host species that enter all of the association results reported here, including Figure 4.

The preceding data about host information can be summarized as an Eltonian shortfall: **252**

(**93**%) of the **272** potential EcM host species native to Canada lack any records of associated EcM fungal partners derived from root material in either GlobalFungi or GenBank.

We next consider the full potential interaction matrix for Canada, which can be considered a form of *metaweb*: the complete set of pairings that are conceivable given the species present, from which any realized interaction network is a subset (Poisot et al. 2015; Strydom et al. 2021). This metaweb represents a logical maximum rather than a claim that every cell is biologically available. The subset of interactions that could be realized cannot currently be estimated because its primary determinants, i.e. where each fungus occurs, and which fungus–host pairings are compatible, are themselves largely unknown.

Starting with the full potential interaction matrix for Canada (**272** host species × **1,411** named EcM fungal species = **383,792** cells), **99.7**% lack data. Of the **1,105** filled cells, **519** (**47.0**%) are supported by a single observation. Coarsening to genus, the matrix collapses to **7,440** cells, of which **96.5**% lack data; of the **259** filled cells, **49** (**18.9**%) rest on a single observation.

Next we consider interactions that are geographically possible, and we define this in multiple ways. The first requires that at least one Canadian record of the fungus falls within the host's modelled range. This yields **108,337** cells (**28.2**% of the original matrix) representing geographically possible interactions, of which **99.0**% remain empty. The second requires only that the fungus and host share an ecoregion: **154,719** cells qualify (**40.3**% of the original matrix), of which **99.3**% are empty. The third, most permissive definition requires merely that the partners share an ecozone: **192,028** cells qualify (**50.0**% of the original matrix), of which **99.4**% are empty. The conclusion that the great majority of potential EcM fungus–host

interactions remain undocumented therefore does not depend on how geographic possibility is defined (full results for all six scales tested, and why we report the full, unrestricted matrix as our primary result rather than any of these geographically restricted versions, are in Supplemental Materials SM1, Section **1.4.7**).

Lastly, Figure 4 provides a geographic perspective of the Eltonian shortfall, simultaneously mapping the estimated richness (within 0.5° grid cells) of EcM host species across Canada and the proportion of species in each cell with at least one associated fungal DNA sequence record in Canada.

The way forward

Across the tree of life, our knowledge and understanding of biodiversity is biased and incomplete due to limitations in the extent of available biodiversity data. Since 2015, summarizing these data gaps under the shortfall framework proposed by Hortal et al. (2015) has proved effective, and we have made substantial progress in overcoming these shortfalls for animals (mostly vertebrates) and plants. For groups like EcM fungi, progress has been slower. In countries like Canada, with vast ecosystems dominated by EcM trees and shrubs, these shortfalls hinder the application of eco-evolutionary knowledge in conservation and land management. Our assessment reveals that current knowledge in Canada is severely limited, characterized by large gaps in available data for EcM fungi and exacerbated by shifting methodological practices, multiple species concepts, and changing technologies. As such, our ability to integrate EcM fungi into conservation strategies to protect biodiversity is lacking. These findings suggest that there is an urgent and broadscale need for additional data and

knowledge on the diversity, distribution, and functioning of EcM fungi across Canadian ecosystems.

From a conservation point of view, the most critical shortfalls to address are the Linnean (taxonomic diversity), Wallacean (geographic distributions), Hutchinsonian (environmental niches), and Eltonian (biotic interactions), especially when it comes to systematically prioritizing land for actions like area-based protection amid ongoing climate change (Eckert et al. 2023; O'Connor et al. 2026; Qin et al. 2026). For example, SDMs are almost certainly going to play an increasingly central role in closing the Wallacean shortfall, and the efforts of Qin et al. (2026) and van Nuland et al. (2024) provide valuable foundations upon which to build. Key to this will be improving model discriminatory power and calibration, which at present is weak for EcM fungal taxa (Qin et al. 2026), presumably because (i) most taxa are represented by comparatively few occurrence records; and / or (ii) the occurrence records that are available likely capture only a small portion of the true realized niches (for symbionts like EcM fungi, this includes the abiotic and biotic niche) (see Hutchinsonian and Eltonian shortfalls).

Fortunately, future data collection to overcome these shortfalls for EcM fungi has the potential to benefit from promising synergies. For example, high-throughput molecular methods to identify EcM fungi from spatially-referenced root samples from known host species can simultaneously contribute taxonomic, geographic, environmental, and species interaction data (Figure 5). Moreover, optimizing sampling programs across specific host species, geographies, and environments could rapidly accrue gains in data coverage (Figure 5). Importantly, generating these data can also advance research priorities (e.g. Qin et al. (2026)) and

immediately benefit conservation planning to reach rapidly approaching targets like Canada's commitment to protecting 30% of land by 2030 (United Nations Convention on Biological Diversity 2022), both of which provide researchers with powerful incentives to expand sampling using more optimized designs.

While we fundamentally need more data on EcM fungi in Canada, solving data limitations alone will not be sufficient to overcome all of the seven shortfalls. This is because, unlike other clades, research on EcM fungi lags behind other fields in producing the open access, standardized, and unified databases that have facilitated broadscale synthesis. For example, while plants and other fungi (like AM fungi) benefit from standardized trait databases (Kattge et al. 2020; Antunes et al. 2025) and integrated, openly licensed data infrastructures (Enquist et al. 2026), general trait frameworks are more difficult to develop for EcM fungi owing to their high functional diversity, though there may be promise in using exploration types (Mansfield et al. 2026). By improving consistency among existing trait data, we can promote broadscale spatial or taxonomic investigations of variation in functional diversity. That said, the existing databases (e.g. GlobalFungi, Global Soil Mycobiome, GenBank) have led to rapid progress in studying EcM fungal diversity at scale, and the numerous, recent contributions from the Society for the Protection of Underground Networks (SPUN; https://www.spun.earth/) and collaborators exemplify these efforts (e.g. Větrovský et al. 2019; Tedersoo et al. 2021; Mikryukov et al. 2023; van Nuland et al. 2024; van Galen et al. 2025; van Nuland et al. 2025; Qin et al. 2026).

Finally, for some shortfalls like the Prestonian (abundance patterns) and Darwinian (evolutionary relationships), Canada is essentially starting from scratch. While daunting, this also

opens the door for optimizing how we collect data from the beginning, learning from efforts made in other regions, and targeting the most pressing aspects or questions pertaining to the different data shortfalls. For example, while studies have explored shifts in EcM fungal abundance following specific disturbances like nitrogen addition (Högberg et al. 2021) or warming (Clemmensen et al. 2006), understanding how EcM fungi respond to fire is also critical in a region like Canada, where single fire seasons can produce burns greater in area than ~40% of European nations (in 2023, over 165 thousand km^2 of forest burned in Canada). Post-fire fungal community surveys often show a reduction in detectable EcM fungi, especially following high severity fire that kills EcM host trees (Dove and Hart 2017; Pulido-Chavez et al. 2021; Packard et al. 2023). But EcM fungi have been shown to persist, dormant in the soil, until new host seedlings begin to establish (Baar et al. 1999; Yamaoka et al. 2025). Better understanding of EcM fungal community distribution, dynamics, dispersal, and persistence is therefore critical for supporting ecosystem recovery and management efforts.

Finally, while we are confident about the data and methods we used to evaluate the seven shortfalls, our study has some important limitations that warrant discussion. First, we based our evaluations exclusively on publicly available and discoverable DNA sequence data from GlobalFungi and GenBank, and did not consider other kinds of data including physical specimen occurrence records available through the Global Biodiversity Information Facility (GBIF). While taxonomic inconsistencies and cryptic diversity render such data less reliable for the present goals, they can nonetheless complement future, molecular-based sampling efforts. For example, Figure S5 shows a map of physical specimen occurrence records, pulled from GBIF, for **142** EcM fungal genera, with records distinguished according to whether the

genus is represented within our Canada-specific DNA sequence datasets. Table S4 shows the **32** genera that currently lack associated DNA sequence data in Canada. This list could be used to target future sampling efforts. Second, our shortfall assessments were not informed in any systematic way by data available solely in the primary research literature; rather, we relied exclusively on curated, discoverable, and publicly accessible data repositories like Mycocosm, BioTime, and FungalRoot. Some shortfalls may therefore be less severe in reality than our limited data suggest. Third, like all studies of this kind, the data acquisition, processing, and analysis pipeline (including bioinformatics pipelines implemented by original sources) involves many decisions, some of which are less well-rationalized than others. This yields a "multiverse" of unique analysis pipelines whose outcomes could vary materially **(Steegen et al. 2016)**. Ideally, one would conduct a "multiverse analysis" that reports the results from all unique pipelines **(Short et al. 2025)**. We are currently exploring that option, but in the interim, for full transparency we provide a thorough "Decisions Log" in the Supplemental Materials SM2 that documents all decisions and their rationales. Lastly, the data presented in this study represent the shortfalls realized at a given point in time. We are in the process of developing a web-based dashboard that will update our shortfall analyses (plus complementary analyses) quarterly - a form of "digital twin" (Visser et al. 2026).

Conclusion

While we are making progress in gathering and synthesizing the data we need to advance biodiversity knowledge and conservation, we still have a long way to go when it comes to understanding the diversity and distribution of EcM fungi in Canada. That said, rapid

progress is possible if we optimize future sampling and take advantage of the lessons learned from other taxa and regions. To that end, our assessment marks an important first step in improving our knowledge and understanding of EcM fungi in Canada and provides the foundation for a coordinated, pan-Canadian sampling effort that can be optimized to address multiple shortfalls simultaneously.

Acknowledgements

This work was directly supported by a Canadian Institute for Ecology & Evolution Working Group grant (2025). We are grateful to the Accelerate Okanagan (Kelowna, BC) for providing collaboration space and logistical support. The research programs of JP, JK, and MF are supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant program. SNK was funded by National Science Foundation awards 2106065, 2338421, 2412561. CQ was supported by a grant from the Paul G. Allen Family Foundation to the Society for the Protection of Underground Networks (SPUN).

Data and code availability

Data and code will be made publicly available through a Borealis archive. In the meantime, a complementary GitHub repository (https://github.com/pitherj/Canada_ecto_shortfalls) hosts all relevant code.

References

Abarenkov K et al. 2024. The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: Sequences, taxa and classifications reconsidered. Nucleic Acids Research, 52(D1): D791–D797. https://doi.org/10.1093/nar/gkad1039.

Abe H, Gan L, Murata M, and Nara K. 2024. Habitat fragmentation strongly restricts gene flow in endangered ectomycorrhizal fungal populations: Evidence from *Rhizopogon Togasawarius* , specific to *Pseudotsuga Japonica* , across the entire distribution range. Molecular Ecology, 33(19): e17533. https://doi.org/10.1111/mec.17533.

Abrego N et al. 2024. Airborne DNA reveals predictable spatial and seasonal dynamics of fungi. Nature. https://doi.org/10.1038/s41586-024-07658-9.

Ackerly DD, and Cornwell WK. 2007. A trait-based approach to community assembly: Partitioning of species trait values into within- and among-community components. Ecology Letters, 10(2): 135–145. https://doi.org/10.1111/j.1461-0248.2006.01006.x.

Aguilar-Trigueros CA et al. 2015. Branching out: Towards a trait-based understanding of fungal ecology. Fungal Biology Reviews, 29(1): 34–41. https://doi.org/10.1016/j.fbr.2015.03.001.

808 Allan JR, Venter O, and Watson JEM. 2017. Temporally inter-comparable maps of terrestrial
809 wilderness and the Last of the Wild. Scientific Data, 4(1): 170187. https://doi.org/10.
810 1038/sdata.2017.187.

811 Allen MF, and Kitajima K. 2013. *In Situ* high-frequency observations of mycorrhizas. New
812 Phytologist, 200(1): 222–228. https://doi.org/10.1111/nph.12363.

813 Anthony MA et al. 2022. Forest tree growth is linked to mycorrhizal fungal composition and
814 function across Europe. The ISME Journal. https://doi.org/10.1038/s41396-021-01159-7.

815 Antunes PM et al. 2025. Enhancing consistency in arbuscular mycorrhizal trait-based research
816 to improve predictions of function. Mycorrhiza, 35(2): 14. https://doi.org/10.1007/s00572-
817 025-01187-7.

818 Averill C, Cates LL, Dietze MC, and Bhatnagar JM. 2019. Spatial vs. Temporal controls
819 over soil fungal community similarity at continental and global scales. The ISME Journal,
820 13(8): 2082–2093. https://doi.org/10.1038/s41396-019-0420-1.

821 Baar J, Horton TR, Kretzer AM, and Bruns TD. 1999. Mycorrhizal colonization of Pinus
822 muricata from resistant propagules after a stand-replacing wildfire. New Phytologist,
823 143(2): 409–418. https://doi.org/10.1046/j.1469-8137.1999.00452.x.

824 Bazzicalupo A et al. 2022. Macrofungal conservation in Canada and target species for

assessment: A starting point. FACETS, 7: 448–463. https://doi.org/10.1139/facets-2021-0180.

Bazzicalupo AL et al. 2020. Fungal heavy metal adaptation through single nucleotide polymorphisms and copy-number variation. Molecular Ecology, 29(21): 4157–4169. https://doi.org/10.1111/mec.15618.

Beaulieu JM, and O'Meara BC. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Systematic Biology, 65(4): 583–601. https://doi.org/10.1093/sysbio/syw022.

Bengtsson-Palme J et al. 2013. Improved software detection and extraction of ITS1 and ITS 2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods in Ecology and Evolution, 4(10): 914–919. https://doi.org/10.1111/2041-210X.12073.

Boshuizen HC, and Te Beest DE. 2023. Pitfalls in the statistical analysis of microbiome amplicon sequencing data. Molecular Ecology Resources, 23(3): 539–548. https://doi.org/10.1111/1755-0998.13730.

Box EO. 1996. Plant functional types and climate at the global scale. Journal of Vegetation Science, 7(3): 309–320. https://doi.org/10.2307/3236274.

842 Boyle BL, Maitner B, Barbosa GC, Rethvick SYB, and Enquist BJ. 2024. Native Species
843 Resolver. Available from https://nsr.biendata.org/.

844 Branco S et al. 2015. Genetic isolation between two recently diverged populations of a
845 symbiotic fungus. Molecular Ecology, 24(11): 2747–2758. https://doi.org/10.1111/mec.
846 13132.

847 Brown JH, Stevens GC, and Kaufman DM. 1996. THE GEOGRAPHIC RANGE: Size, Shape,
848 Boundaries, and Internal Structure. Annual Review of Ecology and Systematics, 27(1):
849 597–623. https://doi.org/10.1146/annurev.ecolsys.27.1.597.

850 Brundrett MC, and Tedersoo L. 2020. Resolving the mycorrhizal status of important northern
851 hemisphere trees. Plant and Soil, 454(1): 3–34. https://doi.org/10.1007/s11104-020-
852 04627-9.

853 Cai L, Giraud T, Zhang N, Begerow D, Cai G, and Shivas RG. 2011. The evolution of species
854 concepts and species recognition criteria in plant pathogenic fungi. Fungal Diversity,
855 50(1): 121–133. https://doi.org/10.1007/s13225-011-0127-8.

856 Cai Q et al. 2024. The evolution of ectomycorrhizal symbiosis and host-plant switches are
857 the main drivers for diversification of Amanitaceae (Agaricales, Basidiomycota). BMC
858 Biology, 22(1): 230. https://doi.org/10.1186/s12915-024-02031-8.

Cameron EK et al. 2018. Global gaps in soil biodiversity data. Nature Ecology and Evolution, 2(7): 1042–1043. https://doi.org/10.1038/s41559-018-0573-8.

Carini P et al. 2020. Effects of Spatial Variability and Relic DNA Removal on the Detection of Temporal Dynamics in Soil Microbial Communities. mBio, 11(1): e02776–19. https://doi.org/10.1128/mBio.02776-19.

Carrillo Y, Stuart EK, and Powell JR. 2025. Necromass of Diverse Root-Associated Fungi Suppresses Decomposition of Native Soil Carbon via Impacts of Their Traits. Ecology Letters, 28(9): e70216. https://doi.org/10.1111/ele.70216.

Cazabonne J, Walker AK, Lesven J, and Haelewaters D. 2024. Singleton-based species names and fungal rarity: Does the number really matter? IMA Fungus, 15(1): 7. https://doi.org/10.1186/s43008-023-00137-2.

Chagnon P-L, Bradley RL, Maherali H, and Klironomos JN. 2013. A trait-based framework to understand life history of mycorrhizal fungi. Trends in Plant Science, 18(9): 484–491. https://doi.org/10.1016/j.tplants.2013.05.001.

Chang Y et al. 2019. Phylogenomics of Endogonaceae and evolution of mycorrhizas within Mucoromycota. New Phytologist, 222(1): 511–525. https://doi.org/10.1111/nph.15613.

Chao A. 1987. Estimating the population size for capture-recapture data with unequal

876 catchability. Biometrics. Journal of the International Biometric Society, 43: 783-791-783-
877 791.

878 Chiu C-H, and Chao A. 2016. Estimating and comparing microbial diversity in the presence
879 of sequencing errors. PeerJ, 4: e1634. https://doi.org/10.7717/peerj.1634.

880 Clemmensen KE, Michelsen A, Jonasson S, and Shaver GR. 2006. Increased ectomycorrhizal
881 fungal abundance after long-term fertilization and warming of two arctic tundra ecosystems.
882 New Phytologist, 171(2): 391–404. https://doi.org/10.1111/j.1469-8137.2006.01778.x.

883 Copot O, Lõhmus A, Abarenkov K, Tedersoo L, and Runnel K. 2024. Contribution of
884 environmental DNA toward fungal Red Listing. Frontiers in Ecology and the Environment,
885 22(9): e2791. https://doi.org/10.1002/fee.2791.

886 Cunningham JA, Padamsee M, Wilson S, and Costello MJ. 2024. Fungi species description
887 rates confirm high global diversity and suggest half remain unnamed. Frontiers of
888 Biogeography, 16(2). https://doi.org/10.21425/F5FBG62358.

889 Dahlberg A, and Mueller GM. 2011. Applying IUCN red-listing criteria for assessing and
890 reporting on the conservation status of fungal species. Fungal Ecology, 4(2): 147–162.
891 https://doi.org/10.1016/j.funeco.2010.11.001.

892 Daru BH. 2024. Predicting undetected native vascular plant diversity at a global scale.

Proceedings of the National Academy of Sciences, 121(34): e2319989121. https://doi.org/10.1073/pnas.2319989121.

Daru BH. 2025. A global database of butterfly species native distributions. Ecology, 106(1): e4462. https://doi.org/10.1002/ecy.4462.

Dauphin B, and Peter M. 2023. Advancing research on ectomycorrhizal fungal adaptation with landscape genomics. Trends in Microbiology, 31(5): 439–443. https://doi.org/10.1016/j.tim.2023.02.002.

Dauphin B, and Peter M. 2024. Tracking signatures of selection in natural populations of ectomycorrhizal fungi – progress, challenges, and prospects. New Phytologist, 242(2): 384–388. https://doi.org/10.1111/nph.19553.

Davison J et al. 2021. Temperature and pH define the realised niche space of arbuscular mycorrhizal fungi. New Phytologist, 231(2): 763–776. https://doi.org/10.1111/nph.17240.

Defrenne CE et al. 2021. High-resolution minirhizotrons advance our understanding of root-fungal dynamics in an experimentally warmed peatland. PLANTS, PEOPLE, PLANET, 3(5): 640–652. https://doi.org/10.1002/ppp3.10172.

Delhaye G, Van Der Linde S, Bauman D, Orme CDL, Suz LM, and Bidartondo MI. 2024. Ectomycorrhizal fungi are influenced by ecoregion boundaries across Europe. Global

910 Ecology and Biogeography, 33(6): e13837. https://doi.org/10.1111/geb.13837.

911 Dentinger BTM, Didukh MY, and Moncalvo J-M. 2011. Comparing COI and ITS as DNA
912 Barcode Markers for Mushrooms and Allies (Agaricomycotina). PLoS ONE, 6(9): e25081.
913 https://doi.org/10.1371/journal.pone.0025081.

914 Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO, and Bini LM. 2013. Darwinian shortfalls
915 in biodiversity conservation. Trends in Ecology & Evolution, 28(12): 689–695. https:
916 //doi.org/10.1016/j.tree.2013.09.003.

917 Dornelas M et al. 2025. BioTIME 2.0: Expanding and Improving a Database of Biodiversity
918 Time Series. Global Ecology and Biogeography, 34(5): e70003. https://doi.org/10.1111/
919 geb.70003.

920 Douhan GW, and Rizzo DM. 2005. Phylogenetic divergence in a local population of the
921 ectomycorrhizal fungus *Cenococcum Geophilum*. New Phytologist, 166(1): 263–271.
922 https://doi.org/10.1111/j.1469-8137.2004.01305.x.

923 Dove NC, and Hart SC. 2017. Fire reduces fungal species richness and in situ mycorrhizal
924 colonization: A meta-analysis. Fire Ecology, 13(2): 37–65. https://doi.org/10.4996/
925 fireecology.130237746.

926 Eckert I, Brown A, Caron D, Riva F, and Pollock LJ. 2023. 30×30 biodiversity gains rely on

national coordination. Nature Communications, 14(1): 7113. https://doi.org/10.1038/s41467-023-42737-x.

Edwards CB et al. 2025. Rapid butterfly declines across the United States during the 21st century. Science, 387(6738): 1090–1094. https://doi.org/10.1126/science.adp4671.

Ekblad A, Wallander H, and Näsholm T. 1998. Chitin and ergosterol combined to measure total and living fungal biomass in ectomycorrhizas. New Phytologist, 138(1): 143–149. https://doi.org/10.1046/j.1469-8137.1998.00891.x.

Elena Schmitz J, and Rahmann S. 2025. A comprehensive review and evaluation of species richness estimation. Briefings in Bioinformatics, 26(2): bbaf158. https://doi.org/10.1093/bib/bbaf158.

Enquist BJ et al. 2026. BIEN: A biodiversity informatics ecosystem advancing open and reproducible workflows for plant observation, plot and trait data. Methods in Ecology and Evolution, 17(5): 1556–1584. https://doi.org/10.1111/2041-210X.70274.

Fry EL et al. 2019. Using plant, microbe, and soil fauna traits to improve the predictive power of biogeochemical models. Methods in Ecology and Evolution, 10(1): 146–157. https://doi.org/10.1111/2041-210X.13092.

García-Roselló E, González-Dacosta J, and Lobo JM. 2023. The biased distribution of

existing information on biodiversity hinders its use in conservation, and we need an integrative approach to act urgently. Biological Conservation, 283: 110118. https://doi.org/10.1016/j.biocon.2023.110118.

Geml J, Morgado LN, Semenova TA, Welker JM, Walker MD, and Smets E. 2015. Long-term warming alters richness and composition of taxonomic and functional groups of arctic fungi. FEMS Microbiology Ecology, 91(8): fiv095. https://doi.org/10.1093/femsec/fiv095.

Golan J et al. 2024. Death caps (*Amanita Phalloides*) frequently establish from sexual spores, but individuals can grow large and live for more than a decade in invaded forests. New Phytologist, 242(4): 1753–1770. https://doi.org/10.1111/nph.19483.

Green JL, Bohannan BJM, and Whitaker RJ. 2008. Microbial biogeography: From taxonomy to traits. Science (New York, N.Y.), 320(5879): 1039–43. https://doi.org/10.1126/science.1153475.

Grigoriev IV et al. 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic acids research, 42(D1): D699–D704.

Hao X, Holyoak M, Zhang Z, and Yan C. 2025. Global Projection of Terrestrial Vertebrate Food Webs Under Future Climate and Land-Use Changes. Global Change Biology, 31(2): e70061. https://doi.org/10.1111/gcb.70061.

961 Hastie T, and Fithian W. 2013. Inference from presence-only data; the ongoing controversy.
962 Ecography, 36(8): 864–867. https://doi.org/10.1111/j.1600-0587.2013.00321.x.

963 Hawkins H-J et al. 2023. Mycorrhizal mycelium as a global carbon pool. Current Biology,
964 33(11): R560–R573. https://doi.org/10.1016/j.cub.2023.02.027.

965 Hébert K et al. 2026. Blitz the Gap: A nation-wide effort to guide citizen science toward the
966 needs of biodiversity science.

967 Hibbett DS, Gilbert L-B, and Donoghue MJ. 2000. Evolutionary instability of ectomycorrhizal
968 symbioses in basidiomycetes. Nature, 407(6803): 506–508. https://doi.org/10.1038/
969 35035065.

970 Hochkirch A et al. 2021. A strategy for the next decade to address data deficiency in neglected
971 biodiversity. Conservation Biology, 35(2): 502–509. https://doi.org/10.1111/cobi.13589.

972 Högberg MN, Högberg P, Wallander H, and Nilsson L-O. 2021. Carbon–nitrogen relations of
973 ectomycorrhizal mycelium across a natural nitrogen supply gradient in boreal forest. New
974 Phytologist, 232(4): 1839–1848. https://doi.org/10.1111/nph.17701.

975 Hortal J, De Bello F, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, and Ladle RJ. 2015. Seven
976 Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annual Review of Ecology,
977 Evolution, and Systematics, 46(1): 523–549. https://doi.org/10.1146/annurev-ecolsys-

112414-054400.

Hudson LN et al. 2017. The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. Ecology and Evolution, 7(1): 145–188. https://doi.org/10.1002/ece3.2579.

IUCN. 2025. IUCN Red List.

James TY, Stajich JE, Hittinger CT, and Rokas A. 2020. Toward a Fully Resolved Fungal Tree of Life. Annual Review of Microbiology, 74(Volume 74, 2020): 291–313. https://doi.org/10.1146/annurev-micro-022020-051835.

Janowski D, and Leski T. 2023. Methods for identifying and measuring the diversity of ectomycorrhizal fungi. Forestry: An International Journal of Forest Research, 96(5): 639–652. https://doi.org/10.1093/forestry/cpad017.

Jörgensen K, Clemmensen KE, Fransson P, Manzoni S, Wallander H, and Lindahl BD. 2025. A trait spectrum linking nitrogen acquisition and carbon use of ectomycorrhizal fungi. New Phytologist, 246(6): 2425–2434. https://doi.org/10.1111/nph.70129.

Karachaliou E, Schmidt C, De Greef E, Docker MF, and Garroway CJ. 2025. Urbanisation Is Associated With Reduced Genetic Diversity in Marine Fish Populations. Molecular Ecology, 34(7): e17711. https://doi.org/10.1111/mec.17711.

995 Kardol P, Throop HL, Adkins J, and Graaff M-AD. 2016. A hierarchical framework for
996 studying the role of biodiversity in soil food web processes and ecosystem services. Soil
997 Biology and Biochemistry, 102: 33–36. https://doi.org/10.1016/j.soilbio.2016.05.002.

998 Kattge J et al. 2020. TRY plant trait database – enhanced coverage and open access. Global
999 Change Biology, 26(1): 119–188. https://doi.org/10.1111/gcb.14904.

1000 Kawahara AY et al. 2023. A global phylogeny of butterflies reveals their evolutionary history,
1001 ancestral hosts and biogeographic origins. Nature Ecology & Evolution, 7(6): 903–913.
1002 https://doi.org/10.1038/s41559-023-02041-9.

1003 Kivlin SN, and Hawkes CV. 2020. Spatial and temporal turnover of soil microbial communities
1004 is not linked to function in a primary tropical forest. Ecology, 101(4): e02985. https:
1005 //doi.org/10.1002/ecy.2985.

1006 Kivlin SN, Hawkes CV, Papeş M, Treseder KK, and Averill C. 2021. The future of microbial
1007 ecological niche theory and modeling. New Phytologist, 231(2): 508–511. https://doi.
1008 org/10.1111/nph.17373.

1009 Kobayashi Y et al. 2023. The genome of Lyophyllum shimeji provides insight into the
1010 initial evolution of ectomycorrhizal fungal genomes. DNA Research, 30(1): dsac053.
1011 https://doi.org/10.1093/dnares/dsac053.

1012 Koch BJ et al. 2018. Estimating taxon-specific population dynamics in diverse microbial
1013 communities. Ecosphere, 9(1): e02090. https://doi.org/10.1002/ecs2.2090.

1014 Kranabetter JM, Berch SM, MacKinnon JA, Ceska O, Dunn DE, and Ott PK. 2018. Species–
1015 area curve and distance–decay relationships indicate habitat thresholds of ectomycorrhizal
1016 fungi in an old-growth Pseudotsuga menziesii landscape. Diversity and Distributions,
1017 24(6): 755–764. https://doi.org/10.1111/ddi.12720.

1018 Kranabetter JM, Dyer T, McPolin MC, Rianhard OM, Philpott TJ, and Beese WJ. 2026.
1019 Spore deposition by ectomycorrhizal and polyporous fungi depicts timber harvesting
1020 impacts and retention patch effectiveness in a coastal British Columbia montane forest.
1021 Biodiversity and Conservation, 35(2): 67. https://doi.org/10.1007/s10531-026-03276-1.

1022 Kwon-Chung KJ et al. 2017. The Case for Adopting the "Species Complex" Nomenclature
1023 for the Etiologic Agents of Cryptococcosis. mSphere, 2(1): e00357–16. https://doi.org/
1024 10.1128/mSphere.00357-16.

1025 Lebreton A et al. 2022. Comparative genomics reveals a dynamic genome evolution in the
1026 ectomycorrhizal milk-cap (*Lactarius*) mushrooms. New Phytologist, 235(1): 306–319.
1027 https://doi.org/10.1111/nph.18143.

1028 Ledger SEH et al. 2023. Past, present, and future of the Living Planet Index. npj Biodiversity,
1029 2(1): 12. https://doi.org/10.1038/s44185-023-00017-3.

1030 Li GC, Lambers H, and Veresoglou SD. 2023. Mycorrhizal ecology would benefit from
1031 region-specific hypotheses. Pedobiologia, 101: 150908. https://doi.org/10.1016/j.pedobi.
1032 2023.150908.

1033 Li J, and Prentice IC. 2024. Global patterns of plant functional traits and their relationships
1034 to climate. Communications Biology, 7(1): 1136. https://doi.org/10.1038/s42003-024-
1035 06777-3.

1036 Li Y et al. 2021. A genome-scale phylogeny of the kingdom Fungi. Current Biology, 31(8):
1037 1653–1665.e5. https://doi.org/10.1016/j.cub.2021.01.074.

1038 Lindahl BD, Clemmensen KE, Stendahl J, and Dahlberg A. 2026. Long-term effects of clear-
1039 cutting forestry on ectomycorrhizal fungi in boreal forest. New Phytologist, nph.71165.
1040 https://doi.org/10.1111/nph.71165.

1041 Lofgren L et al. 2024. Suillus: An emerging model for the study of ectomycorrhizal ecology
1042 and evolution. New Phytologist, 242(4): 1448–1475. https://doi.org/10.1111/nph.19700.

1043 Lomolino MV. 2004. Conservation biogeography. *In* Frontiers of Biogeography: New
1044 Directions in the Geography of Nature. Vol. 293. Sinauer Associates.

1045 Looney BP, Ryberg M, Hampe F, Sánchez-García M, and Matheny PB. 2016. Into and out
1046 of the tropics: Global diversification patterns in a hyperdiverse clade of ectomycorrhizal

fungi. Molecular Ecology, 25(2): 630–647. https://doi.org/10.1111/mec.13506.

Lumbierres M et al. 2022. Area of Habitat maps for the world's terrestrial birds and mammals. Scientific Data, 9(1): 749. https://doi.org/10.1038/s41597-022-01838-w.

Maitner BS et al. 2018. The BIEN R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. Methods in Ecology and Evolution, 9(2): 373–379. https://doi.org/10.1111/2041-210X.12861.

Malloch DW, Pirozynski KA, and Raven PH. 1980. Ecological and evolutionary significance of mycorrhizal symbioses in vascular plants (A Review). Proceedings of the National Academy of Sciences, 77(4): 2113–2118. https://doi.org/10.1073/pnas.77.4.2113.

Mansfield TM et al. 2026. Ecological and genomic variation in ectomycorrhizal fungal exploration types. New Phytologist, nph.71157. https://doi.org/10.1111/nph.71157.

Martin F, Kohler A, Murat C, Veneault-Fourrey C, and Hibbett DS. 2016. Unearthing the roots of ectomycorrhizal symbioses. Nature Reviews Microbiology, 14(12): 760–773. https://doi.org/10.1038/nrmicro.2016.149.

Martin FM, and Van Der Heijden MGA. 2024. The mycorrhizal symbiosis: Research frontiers in genomics, ecology, and agricultural application. New Phytologist, nph.19541. https://doi.org/10.1111/nph.19541.

Martinović T et al. 2021. Temporal turnover of the soil microbiome composition is guild-specific. Ecology Letters, 24(12): 2726–2738. https://doi.org/10.1111/ele.13896.

Martiny AC, Treseder K, and Pusch G. 2013. Phylogenetic conservatism of functional traits in microorganisms. The ISME Journal, 7(4): 830–838. https://doi.org/10.1038/ismej.2012.160.

Matheny PB et al. 2009. Out of the Palaeotropics? Historical biogeography and diversification of the cosmopolitan ectomycorrhizal mushroom family Inocybaceae. Journal of Biogeography, 36(4): 577–592. https://doi.org/10.1111/j.1365-2699.2008.02055.x.

Matias Rodrigues JF et al. 2026. The MicrobeAtlas database: Global trends and insights into Earth's microbial ecosystems. Cell, S009286742600108X. https://doi.org/10.1016/j.cell.2026.01.021.

Mikryukov V et al. 2023. Connecting the multiple dimensions of global soil fungal diversity. Science Advances, 9(48): eadj8016. https://doi.org/10.1126/sciadv.adj8016.

Miralles A et al. 2020. Repositories for Taxonomic Data: Where We Are and What is Missing. Systematic Biology, 69(6): 1231–1253. https://doi.org/10.1093/sysbio/syaa026.

Molina R, and Horton TR. 2015. Mycorrhiza Specificity: Its Role in the Development and Function of Common Mycelial Networks. *In* Mycorrhizal Networks. pp. 1–39.

Moudrý V et al. 2024. Optimising occurrence data in species distribution models: Sample size, positional uncertainty, and sampling bias matter. Ecography, 2024(12): e07294. https://doi.org/10.1111/ecog.07294.

Moulatlet GM et al. 2025. General laws of biodiversity: Climatic niches predict plant range size and ecological dominance globally. Proceedings of the National Academy of Sciences, 122(46): e2517585122. https://doi.org/10.1073/pnas.2517585122.

Mueller GM et al. 2022. What Do the First 597 Global Fungal Red List Assessments Tell Us about the Threat Status of Fungi? Diversity, 14(9): 736. https://doi.org/10.3390/d14090736.

Natural Resources Canada. 2022. The State of Canada's Forests: Annual Report.

Nguyen NH, Hynson NA, and Bruns TD. 2012. Stayin' alive: Survival of mycorrhizal fungal propagules from 6-yr-old forest soil. Fungal Ecology, 5(6): 741–746. https://doi.org/10.1016/j.funeco.2012.05.006.

Nilsson RH et al. 2019. The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research, 47(D259–D264): 259–264. https://doi.org/10.1093/nar/gky1022.

Ning C, Egerton-Warburton LM, Mueller GM, Xiang W, Yan W, and Liu S. 2021. Shifts in

ectomycorrhizal fungal community composition during the early establishment of native and exotic pine seedlings. Applied Soil Ecology, 157: 103722. https://doi.org/10.1016/j.apsoil.2020.103722.

Niskanen T et al. 2023. Pushing the Frontiers of Biodiversity Research: Unveiling the Global Diversity, Distribution, and Conservation of Fungi. Annual Review of Environment and Resources, 48(1): annurev-environ-112621-090937. https://doi.org/10.1146/annurev-environ-112621-090937.

Noffsinger CR et al. 2026. Biogeography and host associations of *Russula* subsection *Xerampelinae* based on large-scale analysis of UNITE sequence data. New Phytologist, 249(5): 2547–2558. https://doi.org/10.1111/nph.70833.

Nuñez MA, and Dickie IA. 2014. Invasive belowground mutualists of woody plants. Biological Invasions, 16(3): 645–661. https://doi.org/10.1007/s10530-013-0612-y.

Nuñez MA, Horton TR, and Simberloff D. 2009. Lack of belowground mutualisms hinders Pinaceae invasions. Ecology, 90(9): 2352–2359. https://doi.org/10.1890/08-2139.1.

O'Connor LMJ et al. 2026. The Untapped Potential of Food Webs in Systematic Conservation Planning. Conservation Letters, 19(1): e70002. https://doi.org/10.1111/con4.70002.

Olsson PA. 1999. Signature fatty acids provide tools for determination of the distribution

and interactions of mycorrhizal fungi in soil. FEMS Microbiology Ecology, 29(4): 303–310. https://doi.org/10.1111/j.1574-6941.1999.tb00621.x.

Olsson PA, Münzenberger B, Mahmood S, and Erland S. 2000. Molecular and anatomical evidence for a three-way association between Pinus sylvestris and the ectomycorrhizal fungi Suillus bovinus and Gomphidius roseus. Mycological Research, 104(11): 1372–1378. https://doi.org/10.1017/S0953756200002823.

Ovaskainen O et al. 2024. Global Spore Sampling Project: A global, standardized dataset of airborne fungal DNA. Scientific Data, 11(1): 561. https://doi.org/10.1038/s41597-024-03410-0.

Packard EE, Durall DM, and Jones MD. 2023. Successional changes in fungal communities occur a few weeks following wildfire in a mixed Douglas-fir-ponderosa pine forest. Fungal Ecology, 63: 101246. https://doi.org/10.1016/j.funeco.2023.101246.

Parisy B et al. 2025. Opportunistic partner choice among arctic plants and root-associated fungi is driven by environmental conditions. Ecological Monographs, 95(3): e70038. https://doi.org/10.1002/ecm.70038.

Peay KG, Schubert MG, Nguyen NH, and Bruns TD. 2012. Measuring ectomycorrhizal fungal dispersal: Macroecological patterns driven by microscopic propagules. Molecular Ecology, 21(16): 4122–4136. https://doi.org/10.1111/j.1365-294X.2012.05666.x.

Peter M et al. 2016. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus Cenococcum geophilum. Nature Communications, 7(1): 12662. https://doi.org/10.1038/ncomms12662.

Pither J, and Pickles BJ. 2017. The paleosymbiosis hypothesis: Host plants can be colonised by root symbionts that have been inactive for centuries to millenia. FEMS Microbiology Ecology, 93(6): fix061. https://doi.org/10.1093/femsec/fix061.

Poelen JH, Simons JD, and Mungall CJ. 2014. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. Ecological Informatics, 24: 148–159. https://doi.org/10.1016/j.ecoinf.2014.08.005.

Poisot T, Stouffer DB, and Gravel D. 2015. Beyond species: Why ecological interaction networks vary through space and time. Oikos, 124(3): 243–251. https://doi.org/10.1111/oik.01719.

Policelli N, Hoeksema JD, Moyano J, Vilgalys R, Vivelo S, and Bhatnagar JM. 2023. Global pine tree invasions are linked to invasive root symbionts. New Phytologist, 237(1): 16–21. https://doi.org/10.1111/nph.18527.

Põlme S et al. 2020. FungalTraits: A user-friendly traits database of fungi and fungus-like stramenopiles. Fungal Diversity, 105(1): 1–16. https://doi.org/10.1007/s13225-020-00466-2.

Pulido-Chavez MF, Alvarado EC, DeLuca TH, Edmonds RL, and Glassman SI. 2021. High-severity wildfire reduces richness and alters composition of ectomycorrhizal fungi in low-severity adapted ponderosa pine forests. Forest Ecology and Management, 485: 118923. https://doi.org/10.1016/j.foreco.2021.118923.

Qin C et al. 2026. High-Resolution Range Mapping of Mycorrhizal Fungal Species Reveals Systematic Biases in Their Protection. Conservation Letters, 19(3): e70051. https://doi.org/10.1111/con4.70051.

Qin C, Pellitier PT, Van Nuland ME, Peay KG, and Zhu K. 2023. Niche modelling predicts that soil fungi occupy a precarious climate in boreal forests. Global Ecology and Biogeography, 32(7): 1127–1139. https://doi.org/10.1111/geb.13684.

Quinn TP, Erb I, Gloor G, Notredame C, Richardson MF, and Crowley TM. 2019. A field guide for the compositional analysis of any-omics data. GigaScience, 8(9): giz107. https://doi.org/10.1093/gigascience/giz107.

Read DJ. 1991. Mycorrhizas in ecosystems. Experientia, 47(4): 376–391. https://doi.org/10.1007/BF01972080.

Richter DL, and Bruhn JN. 1989. Pinus resinosa ectomycorrhizae: Seven host-fungus combinations synthesized in pure culture. 7(3): 211–228.

Rognes T, Flouri T, Nichols B, Quince C, and Mahé F. 2016. VSEARCH: A versatile open source tool for metagenomics. PeerJ, 4: e2584. https://doi.org/10.7717/peerj.2584.

Ryberg M, and Matheny PB. 2011. Asynchronous origins of ectomycorrhizal clades of Agaricales. Proceedings of the Royal Society B: Biological Sciences, 279(1735): 2003–2011. https://doi.org/10.1098/rspb.2011.2428.

Sánchez-García M, Matheny PB, Palfner G, and Lodge DJ. 2014. Deconstructing the Tricholomataceae (Agaricales) and introduction of the new genera Albomagister, Corneriella, Pogonoloma and Pseudotricholoma. TAXON, 63(5): 993–1007. https://doi.org/10.12705/635.635.3.

Sánchez-García M, Ryberg M, Khan FK, Varga T, Nagy LG, and Hibbett DS. 2020. Fruiting body form, not nutritional mode, is the major driver of diversification in mushroom-forming fungi. Proceedings of the National Academy of Sciences, 117(51): 32528–32534. https://doi.org/10.1073/pnas.1922539117.

Sato H. 2024. The evolution of ectomycorrhizal symbiosis in the Late Cretaceous is a key driver of explosive diversification in Agaricomycetes. New Phytologist, 241(1): 444–460. https://doi.org/10.1111/nph.19055.

Sato H, Tanabe AS, and Toju H. 2017. Host shifts enhance diversification of ectomycorrhizal fungi: Diversification rate analysis of the ectomycorrhizal fungal genera Strobilomyces

and Afroboletus with an 80-gene phylogeny. New Phytologist, 214(1): 443–454. https://doi.org/10.1111/nph.14368.

Sato H, and Toju H. 2019. Timing of evolutionary innovation: Scenarios of evolutionary diversification in a species-rich fungal clade, Boletales. New Phytologist, 222(4): 1924–1935. https://doi.org/10.1111/nph.15698.

Schmidt C, Domaratzki M, Kinnunen RP, Bowman J, and Garroway CJ. 2020. Continent-wide effects of urbanization on bird and mammal genetic diversity. Proceedings of the Royal Society B: Biological Sciences, 287(1920): 20192497. https://doi.org/10.1098/rspb.2019.2497.

Shaw RE et al. 2025. Global meta-analysis shows action is needed to halt genetic diversity loss. Nature, 638(8051): 704–710. https://doi.org/10.1038/s41586-024-08458-x.

Shelton BR et al. 2026. Fungal ecology in the age of 'omics. New Phytologist, 250(2): 822–833. https://doi.org/10.1111/nph.70900.

Shi X, Zhang S, Mueller GM, Liu P, Yu F, and Senanayake IC. 2025. A subgeneric revision of the genus Suillus (Suillaceae, Boletales) and novel taxa from Eastern Asia based on morphology and multigene phylogenies. IMA Fungus, 16: e144260. https://doi.org/10.3897/imafungus.16.144260.

Short CA, Inceler YC, Frank M, and Hildebrandt A. 2025. The Systematic Multiverse Analysis Registration Tool for defining multiverse analyses. Royal Society Open Science, 12(10): 250800. https://doi.org/10.1098/rsos.250800.

Smith SA, and Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. American Journal of Botany, 105(3): 302–314. https://doi.org/10.1002/ajb2.1019.

Sønstebø JH et al. 2022. Population genomics of a forest fungus reveals high gene flow and climate adaptation signatures. Molecular Ecology, 31(7): 1963–1979. https://doi.org/10.1111/mec.16369.

Sothe C, Gonsamo A, Arabian J, Kurz WA, Finkelstein SA, and Snider J. 2022. Large Soil Carbon Storage in Terrestrial Ecosystems of Canada. Global Biogeochemical Cycles, 36(2): e2021GB007213. https://doi.org/10.1029/2021GB007213.

Soudzilovskaia NA et al. 2015. Global patterns of plant root colonization intensity by mycorrhizal fungi explained by climate and soil chemistry. Global Ecology and Biogeography, 24(3): 371–382. https://doi.org/10.1111/geb.12272.

Soudzilovskaia NA et al. 2020. FungalRoot: Global online database of plant mycorrhizal associations. New Phytologist, 227(3): 955–966. https://doi.org/10.1111/nph.16569.

Steegen S, Tuerlinckx F, Gelman A, and Vanpaemel W. 2016. Increasing Transparency

Through a Multiverse Analysis. Perspectives on Psychological Science, 11(5): 702–712. https://doi.org/10.1177/1745691616658637.

Steidinger BS et al. 2019. Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. Nature, 569(7756): 404–408. https://doi.org/10.1038/s41586-019-1128-0.

Steidinger BS et al. 2020. Ectomycorrhizal fungal diversity predicted to substantially decline due to climate changes in North American Pinaceae forests. Journal of Biogeography, 47(3): 772–782. https://doi.org/10.1111/jbi.13802.

Steinke D, Gemeinholzer B, Martínez-Meyer E, Noesgaard D, Young A, and Schigel D. 2025. Globally aggregated biodiversity data impact predictive and descriptive research. Proceedings of the National Academy of Sciences, 122(50): e2519119122. https://doi.org/10.1073/pnas.2519119122.

Strullu-Derrien C, Selosse M-A, Kenrick P, and Martin FM. 2018. The origin and evolution of mycorrhizal symbioses: From palaeomycology to phylogenomics. The New Phytologist, 220(4): 1012–1030. https://doi.org/10.1111/nph.15076.

Strydom T et al. 2021. A roadmap towards predicting species interaction networks (across space and time). Philosophical Transactions of the Royal Society B: Biological Sciences, 376(1837): 20210063. https://doi.org/10.1098/rstb.2021.0063.

1238 Suenaga H. 2015. Targeted metagenomics unveils the molecular basis for adaptive evolution
1239 of enzymes to their environment. Frontiers in Microbiology, 6. https://doi.org/10.3389/
1240 fmicb.2015.01018.

1241 Talbot JM et al. 2014. Endemism and functional convergence across the North American
1242 soil mycobiome. Proceedings of the National Academy of Sciences, 111(17): 6341–6346.
1243 https://doi.org/10.1073/pnas.1402584111.

1244 Taylor AFS, Martin F, and Read DJ. 2000. Fungal Diversity in Ectomycorrhizal Commu-
1245 nities of Norway Spruce [Picea abies (L.) Karst.] and Beech (Fagus sylvatica L.) Along
1246 North-South Transects in Europe. *In* Carbon and Nitrogen Cycling in European Forest
1247 Ecosystems. Edited by MM Caldwell et al. Vol. 142. Springer Berlin Heidelberg, Berlin,
1248 Heidelberg. pp. 343–365.

1249 Taylor DL et al. 2016. Accurate Estimation of Fungal Diversity and Abundance through
1250 Improved Lineage-Specific Primers Optimized for Illumina Amplicon Sequencing. Applied
1251 and Environmental Microbiology, 82(24): 7217–7226. https://doi.org/10.1128/AEM.
1252 02576-16.

1253 Taylor DL, Hollingsworth TN, McFarland JW, Lennon NJ, Nusbaum C, and Ruess RW. 2014.
1254 A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche
1255 partitioning. Ecological Monographs, 84(1): 3–20. https://doi.org/10.1890/12-1693.1.

1256 Tedersoo L et al. 2012. Towards global patterns in the diversity and community structure of
1257 ectomycorrhizal fungi. Molecular Ecology, 21(17): 4160–4170. https://doi.org/10.1111/j.
1258 1365-294X.2012.05602.x.

1259 Tedersoo L et al. 2014. Global diversity and geography of soil fungi. Science, 346(6213):
1260 1256688. https://doi.org/10.1126/science.1256688.

1261 Tedersoo L et al. 2021. The Global Soil Mycobiome consortium dataset for boosting fungal
1262 diversity research. Fungal Diversity. https://doi.org/10.1007/s13225-021-00493-7.

1263 Tedersoo L et al. 2022. Global patterns in endemicity and vulnerability of soil fungi. Global
1264 Change Biology, 28(22): 6696–6710. https://doi.org/10.1111/gcb.16398.

1265 Tedersoo L, May TW, and Smith ME. 2010. Ectomycorrhizal lifestyle in fungi: Global
1266 diversity, distribution, and evolution of phylogenetic lineages. Mycorrhiza, 20(4): 217–263.
1267 https://doi.org/10.1007/s00572-009-0274-x.

1268 Tedersoo L, and Smith ME. 2013. Lineages of ectomycorrhizal fungi revisited: Foraging
1269 strategies and novel lineages revealed by sequences from belowground. Fungal Biology
1270 Reviews, 27(3-4): 83–99. https://doi.org/10.1016/j.fbr.2013.09.001.

1271 Tedersoo L, and Smith ME. 2017. Ectomycorrhizal Fungal Lineages: Detection of Four New
1272 Groups and Notes on Consistent Recognition of Ectomycorrhizal Taxa in High-Throughput

Sequencing Studies. *In* Biogeography of Mycorrhizal Symbiosis. Edited by L Tedersoo. Springer International Publishing, Cham. pp. 125–142.

Teste FP, Jones MD, and Dickie IA. 2020. Dual-mycorrhizal plants: Their ecology and relevance. New Phytologist, 225(5): 1835–1851. https://doi.org/10.1111/nph.16190.

Thorn S et al. 2016. Changes in the dominant assembly mechanism drive species loss caused by declining resources. Ecology Letters, 19(2): 163–170. https://doi.org/10.1111/ele.12548.

Tikhonov G et al. 2020. Joint species distribution modelling with the R-package Hmsc. Methods in Ecology and Evolution, 11(3): 442–447. https://doi.org/10.1111/2041-210X.13345.

Timling I et al. 2012. Distribution and drivers of ectomycorrhizal fungal communities across the North American Arctic. Ecosphere, 3(October): 1–25. https://doi.org/10.1890/Es12-00217.1.

Tomao A, Antonio Bonet J, Castaño C, and de-Miguel S. 2020. How does forest management affect fungal diversity and community composition? Current knowledge and future perspectives for the conservation of forest fungi. Forest Ecology and Management, 457: 117678. https://doi.org/10.1016/j.foreco.2019.117678.

Tóth BB, and Barta Z. 2010. Ecological studies of ectomycorrhizal fungi: An analysis of survey

methods. Fungal Diversity, 45(1): 3–19. https://doi.org/10.1007/s13225-010-0052-2.

Treseder KK, Schimel JP, Garcia MO, and Whiteside MD. 2010. Slow turnover and production of fungal hyphae during a Californian dry season. Soil Biology and Biochemistry, 42(9): 1657–1660. https://doi.org/10.1016/j.soilbio.2010.06.005.

United Nations Convention on Biological Diversity. 2022. Kunming-Montreal Global Biodiversity Framework.

Upham NS, Esselstyn JA, and Jetz W. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLOS Biology, 17(12): e3000494. https://doi.org/10.1371/journal.pbio.3000494.

van der Linde S et al. 2018. Environment and host as large-scale controls of ectomycorrhizal fungi. Nature, 558(7709): 243–248. https://doi.org/10.1038/s41586-018-0189-9.

van Galen LG et al. 2025. The biogeography and conservation of Earth's "dark" ectomycorrhizal fungi. Current Biology, 35(11): R563–R574. https://doi.org/10.1016/j.cub.2025.03.079.

Van Klink R, Bowler DE, Gongalsky KB, Shen M, Swengel SR, and Chase JM. 2024. Disproportionate declines of formerly abundant species underlie insect loss. Nature, 628(8007): 359–364. https://doi.org/10.1038/s41586-023-06861-4.

van Nuland ME et al. 2025. Global hotspots of mycorrhizal fungal richness are poorly protected. Nature, 645(8080): 414–422. https://doi.org/10.1038/s41586-025-09277-4.

van Nuland ME, Qin C, Pellitier PT, Zhu K, and Peay KG. 2024. Climate mismatches with ectomycorrhizal fungi contribute to migration lag in North American tree range shifts. Proceedings of the National Academy of Sciences, 121(23): e2308811121. https://doi.org/10.1073/pnas.2308811121.

van Proosdij ASJ, Sosef MSM, Wieringa JJ, and Raes N. 2016. Minimum required number of specimen records to develop accurate species distribution models. Ecography, 39(6): 542–552. https://doi.org/10.1111/ecog.01509.

Varga T et al. 2019. Megaphylogeny resolves global patterns of mushroom evolution. Nature Ecology & Evolution, 3(4): 668–678. https://doi.org/10.1038/s41559-019-0834-1.

Větrovský T et al. 2019. A meta-analysis of global fungal distribution reveals climate-driven patterns. Nature Communications, 10(1): 5142. https://doi.org/10.1038/s41467-019-13164-8.

Větrovský T et al. 2020. GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. Scientific Data, 7(1): 228. https://doi.org/10.1038/s41597-020-0567-7.

Vincenot L, and Selosse M-A. 2017. Population Biology and Ecology of Ectomycorrhizal Fungi. *In* Biogeography of Mycorrhizal Symbiosis. Edited by L Tedersoo. Vol. 230. Springer International Publishing, Cham. pp. 39–59.

Visser ME et al. 2026. Digital twins as a tool for ecosystem research. Trends in Ecology & Evolution, S0169534726001035. https://doi.org/10.1016/j.tree.2026.04.016.

Voller F, Ardanuy A, Taylor AFS, and Johnson D. 2024. Maintenance of host specialisation gradients in ectomycorrhizal symbionts. New Phytologist, 242(4): 1426–1435. https://doi.org/10.1111/nph.19395.

Weiss S et al. 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome, 5(1): 27. https://doi.org/10.1186/s40168-017-0237-y.

White ER. 2019. Minimum Time Required to Detect Population Trends: The Need for Long-Term Monitoring Programs. BioScience, 69(1): 40–46. https://doi.org/10.1093/biosci/biy144.

Wilson AW, Hosaka K, and Mueller GM. 2017. Evolution of ectomycorrhizas as a driver of diversification and biogeographic patterns in the model mycorrhizal mushroom genus *Laccaria*. New Phytologist, 213(4): 1862–1873. https://doi.org/10.1111/nph.14270.

www.inaturalist.org. iNaturalist.

Xu J. 2020. Fungal species concepts in the genomics era. Genome, 63(9): 459–468. https://doi.org/10.1139/gen-2020-0022.

Yamaoka NK, Packard EE, and Jones MD. 2025. Nitrogen accumulation accompanies ectomycorrhiza formation in pine germinants the first growing season after wildfire or clearcutting. Mycorrhiza, 35(5): 58. https://doi.org/10.1007/s00572-025-01229-0.

Zanne AE et al. 2020. Fungal functional ecology: Bringing a trait-based approach to plant-associated fungi. Biological Reviews, 95(2): 409–433. https://doi.org/10.1111/brv.12570.

Tables

Table 1: Summary of data derived from v5 of the GlobalFungi dataset. Summary data are provided for the Canada-specific samples, and for the full GlobalFungi dataset for comparison.

Table 2: Properties of the Canadian GenBank EcM fungal sequence dataset, after ITS2/ITS-both filtering and UNITE SH assignment ($\geq$ 98.5% identity). Taxon counts are over all Canadian GenBank records.

Table 3: Occupancy of georeferenced Canadian EcM fungal taxa (GlobalFungi only; sites defined as 30 arc-second grid cells). For each taxonomic level: number of taxa; the first quartile (Q1), median, third quartile (Q3), and maximum number of occupied cells within Canada; and the number (percentage) of taxa recorded in fewer than 30 cells within Canada and across the worldwide GlobalFungi v5 dataset.

Table 4: Trait coverage for the 124 EcM fungal genera detected in Canada, across the six traits within the FungalTraits database applicable to EcM fungi. Coverage is the percentage of genera with a non-missing value; explicit "unknown" entries are counted as missing.

1363 **Table 1**

Category	Metric	Canada	GlobalFungi-wide	Canada as % of GlobalFungi-wide
Sampling effort	Quality-filtered samples, all fungal taxa	1,567	47,606	3.3%
Sampling effort	Quality-filtered samples with ≥ 1 EcM fungi detection	1,469	34,928	4.2%
Geographic coverage	Unique EcM fungi sampling locations	486	11,651	4.2%
Detection volume	EcM fungi detection records (sample $\times$ SH-code, abundance > 0)	39,867	829,687	4.8%
Detection volume	EcM fungi total read abundance (Σ reads)	8,641,968	154,022,074	5.6%
Taxonomic richness	EcM fungal named species detected	1,002	3,759	26.7%
Taxonomic richness	EcM fungal genera detected	103	225	45.8%
Taxonomic richness	EcM fungal SH codes (UNITE v10) detected	2,062	12,970	15.9%

1364 **Table 2**

Property	Value
EcM fungal sequence records	9,403
Records with coordinates	3,453
EcM fungal SH codes	1,644
SH codes with a UNITE species-level name	1,193
EcM fungal genera	109
EcM fungal named species	1,067
Unique georeferenced sampling locations	1,180

Table 3

Taxonomic level	Taxa	Q1 cells (Canada)	Median cells (Canada)	Q3 cells (Canada)	Max cells (Canada)	< 30 cells in Canada	< 30 cells worldwide
Named species	1,002	2	5	17	280	846 (84%)	351 (35%)
SH codes	2,062	1	3	10	306	1,855 (90%)	969 (47%)

Table 4

Trait	Trait class	Genera documented	Coverage (%)
Host specificity	Ecological/interaction	6 / 124	4.8
Secondary lifestyle	Ecological/interaction	3 / 124	2.4
Endophytic interaction capability	Ecological/interaction	2 / 124	1.6
Fruitbody type	Morphological/structural	124 / 124	100.0
Hymenium type	Morphological/structural	124 / 124	100.0
Mycorrhiza exploration type	Morphological/structural	119 / 124	96.0

Figures

Figure 1: Sampling locations from which EcM fungal records were publicly available from either GlobalFungi v5 (n = 486 unique locations) or GenBank (n = 1,180 unique locations), for a combined total of 1,666 unique locations. Many locations overlap so are not clearly visible.

Figure 2: Frequency distributions of the number of 30 arc-second grid cells (~1 km^2) occupied by EcM fungal taxa, GlobalFungi only. Top row: occupancy within Canada for (a) named species (n = 1,002 species) and (b) SH codes (n = 2,062 SH codes). Bottom row: occupancy across the full global GlobalFungi v5 dataset for the same (c) named species and (d) SH codes.

Figure 3: Available climate space and Canadian EcM fungal sampling coverage, in geographic and climate space (WorldClim v2.1 BIO1 × BIO12, 50 × 50 grid of climate zones). Top row (a–c): geographic maps. Bottom row (d–f): the same in climate space. Left (a, d): available climate frequency; in (d), grid cells outlined in white indicate climates never sampled by GlobalFungi + GenBank combined, i.e. the grey zones in (f). Middle (b, e): GlobalFungi coverage; grey areas denote climates not sampled by any of the 486 GlobalFungi sampling locations. Right (c, f): combined GlobalFungi + GenBank coverage; grey areas denote climates not sampled by any of the 1,666 combined sampling locations. Figure style adapted from Hébert et al. (2026).

Figure 4: Bivariate maps of EcM host species richness (BIEN2 modelled ranges) versus proportion of host species with root-derived EcM fungal molecular detection data in Canada,

shown separately for non-tree (shrub and herbaceous; n = 139 species; top) and tree (n = 88 species; bottom) EcM host species. Counts are the host species with a BIEN2 modelled range overlapping Canada, which is fewer than the full native host list (see Supplemental Materials SM1, Section 1.3.4). Sampling unit: 0.5° grid cell; only cells with predicted EcM host habitat are shown.

Figure 5: Efforts to address current data shortfalls for EcM fungi in Canada can benefit from informed and optimized sampling programs that synergize progress to rapidly improve our taxonomic, geographic, and abiotic knowledge of EcM fungi across many host plant species. Ellipses show how sampling targeting a particular region (Wallacean) or abiotic suite of conditions (Hutchinsonian) can simultaneously address other data shortfalls, including Eltonian and Linnean. For example, the blue ellipse shows how sampling the under-represented climate conditions of the arctic (top-right quadrant) can simultaneously address the Linnean shortfall (top left quadrant). Likewise, the green ellipse shows how filling the large geographic sampling gap across the provinces of Manitoba and Saskatchewan (bottom left quadrant) could simultaneously address both the Hutchinsonian and Eltonian shortfalls (right quadrants).

Figure 1

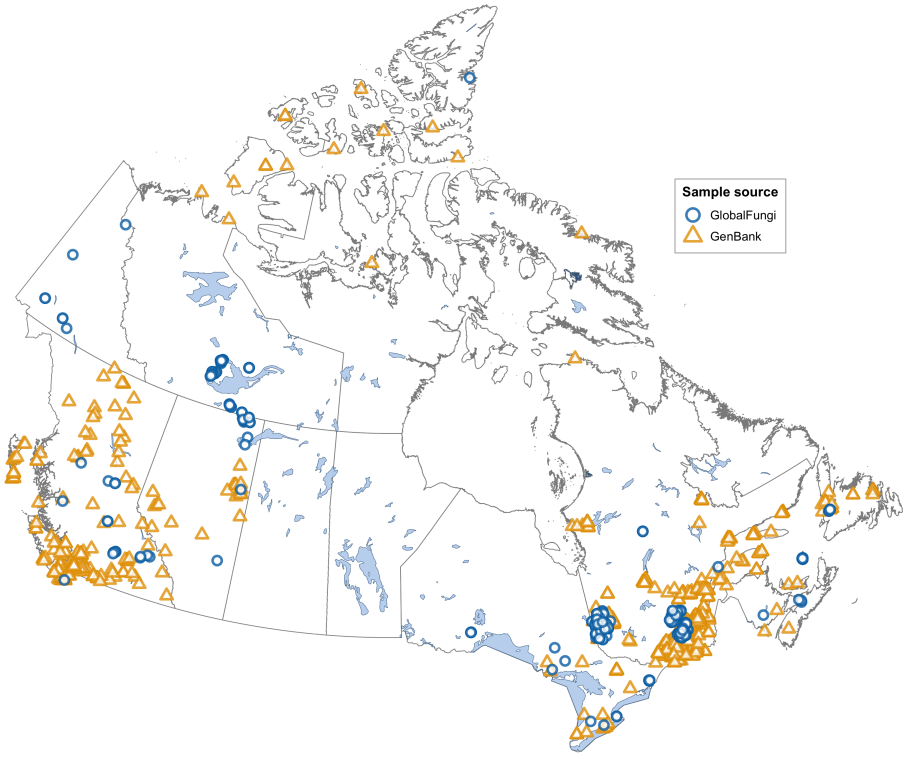
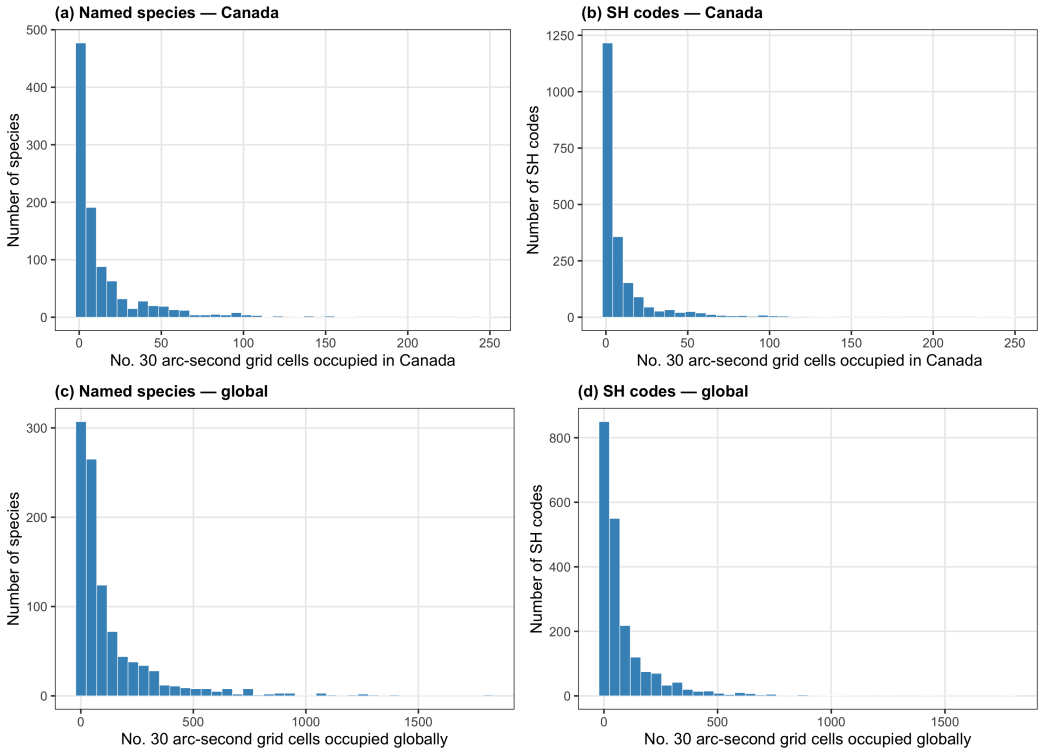
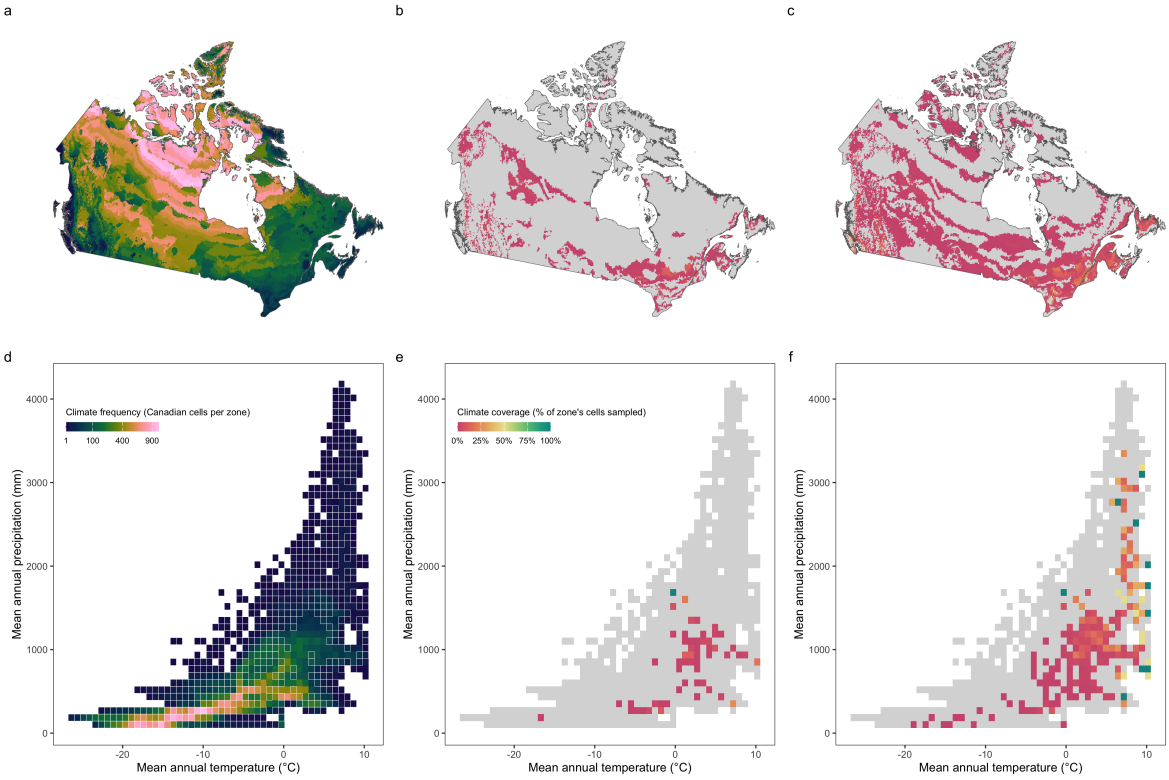


Figure 2


(a) Named species — Canada
Number of species
No. 30 arc-second grid cells occupied in Canada
(b) SH codes — Canada
Number of SH codes
No. 30 arc-second grid cells occupied in Canada
(c) Named species — global
Number of species
No. 30 arc-second grid cells occupied globally
(d) SH codes — global
Number of SH codes
No. 30 arc-second grid cells occupied globally

1407 # Figure 3



1408

Figure 4

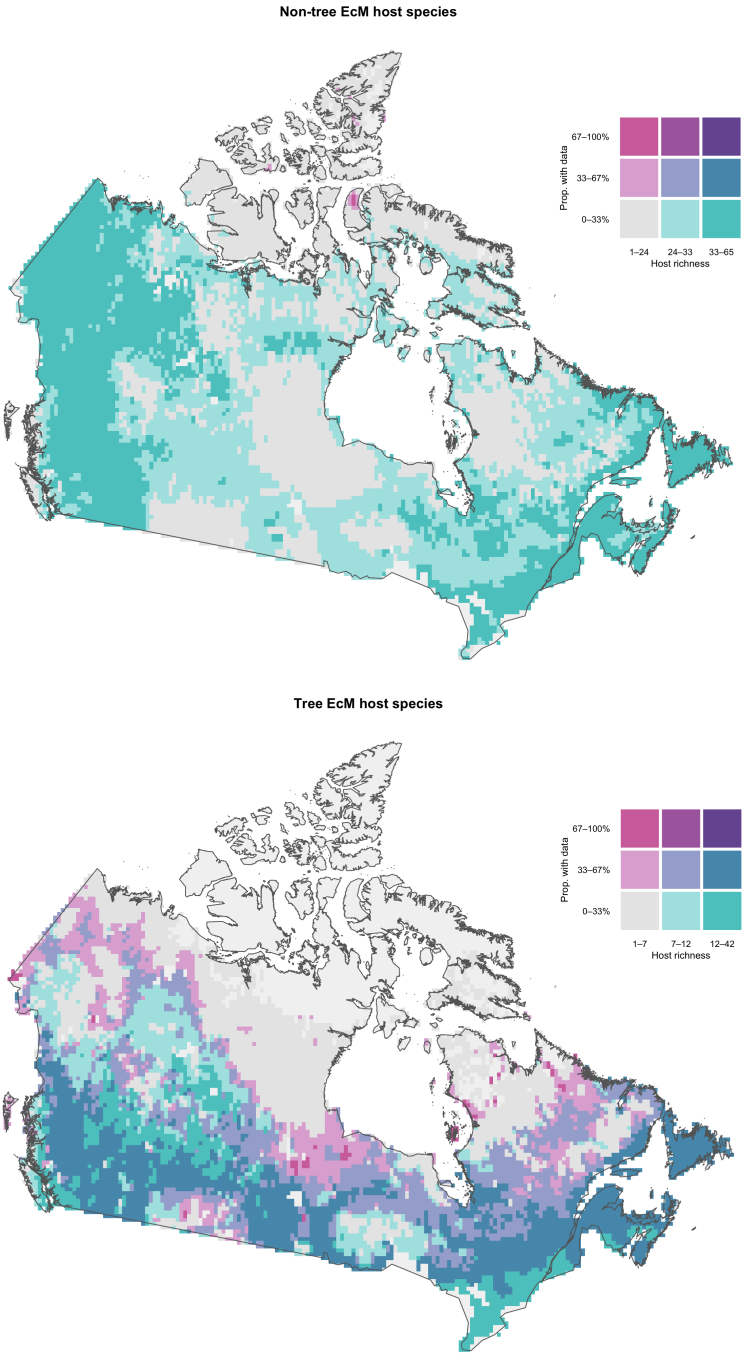
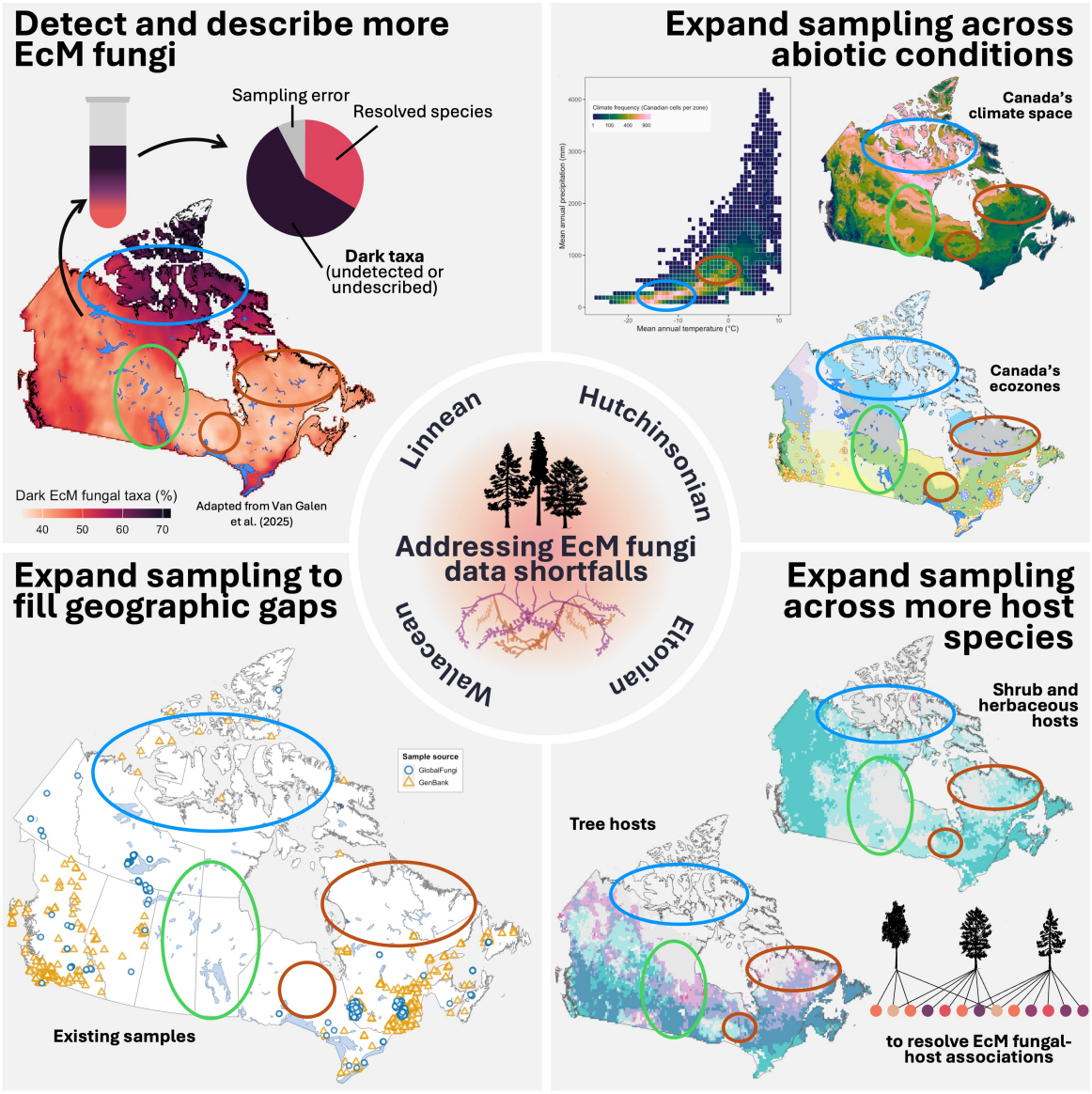

Non-tree EcM host species
Prop. with data
67–100%
33–67%
0–33%
1–24
24–33
33–65
Host richness
Tree EcM host species
Prop. with data
67–100%
33–67%
0–33%
1–7
7–12
12–42
Host richness

Figure 5


Detect and describe more EcM fungi
Sampling error
Resolved species
Dark taxa (undetected or undescribed)
Dark EcM fungal taxa (%)
40 50 60 70
Adapted from Van Galen et al. (2025)
Expand sampling across abiotic conditions
Climate frequency (Canadian cells per zone)
1 100 400 900
Mean annual precipitation (mm)
Mean annual temperature (°C)
Canada's climate space
Canada's ecozones
Linnean
Hutchinsonian
Addressing EcM fungi data shortfalls
Wallacean
Eltonian
Expand sampling to fill geographic gaps
Sample source
GlobalFungi
GenBank
Existing samples
Expand sampling across more host species
Shrub and herbaceous hosts
Tree hosts
to resolve EcM fungal-host associations

Supplemental materials (SM1) for: An assessment of biodiversity data shortfalls for ectomycorrhizal fungi in Canada

Jason Pither, with assistance from Claude (Sonnet 5 and Opus 4.8)

Table of contents

1 Detailed methods

1.1 Software

We conducted analyses in R (v4.5.2; R Core Team (2026)). We used the following R packages: `sf` (Pebesma (2018)) for vector spatial operations; `terra` (Hijmans (2025)) for rasters; `ggplot2` (Wickham (2016)) and `patchwork` (Pedersen (2024)) for figures; `rentrez` (Winter (2017)) for GenBank retrieval; `rgbif` (Chamberlain et al. (2024)) for GBIF access; `BIEN` (Maitner et al. (2018)) for plant range and native-status data; `geodata` (Hijmans (2026)) and `rnaturalearth` (Massicotte and South (2026)) for spatial base layers; `httr` (Wickham (2026)) for programmatic download of BIEN2 range shapefiles; `NSR` (Maitner and Boyle (2024)) for native-status resolution; `GIFT` (Denelle and Weigelt (2024)) for the plant growth-form traits used to fill gaps in the BIEN trait coverage; `readxl` for reading FungalRoot's supplementary workbook; and `data.table` for column-selective reads of the multi-gigabyte GlobalFungi matrices. We used `ITSx` v1.1.3 (Bengtsson-Palme et al. (2013)), with its `HMMER` 3.x backend, to extract ITS sub-regions from GenBank sequences; `vsearch` v2.30.4 (Rognes et al. (2016)) for sequence-similarity searching; and `awk` for extraction from the GlobalFungi abundance matrix.

1.2 Assembly of the Canadian EcM fungal dataset

1.2.1 Data sources

We compiled EcM fungal records for Canada from two publicly accessible sequence sources: version 5 of the GlobalFungi database (Větrovský et al. (2020)) and NCBI GenBank (accessed 29 June 2026). GlobalFungi is a curated, purpose-built database of community-survey amplicon data, consistently processed but aggregated from studies using varied sequencing platforms and protocols. Tallied across the full raw metadata file (`02_globalfungi.R`, Step 2b), soil and soil-derived substrates (soil, topsoil, rhizosphere soil, root + rhizosphere soil) account for **56.3**% of GlobalFungi's **84,972** samples and roots for **18.7**%. Root sampling is therefore not rare in the database as a whole; it is rare in Canada, where roots make up only **4.6**% of samples and **3.0**% of the Canadian samples carrying EcM taxa. The thin root coverage underlying our host analyses is thus a Canadian sampling shortfall rather than a property of the data source. GenBank is a general-purpose sequence repository comprising records deposited from diverse studies and sampling efforts.

1.2.2 UNITE reference database and SH assignment

We assigned fungal species hypothesis (SH) codes and taxonomy using the UNITE general FASTA release (build 4 April 2024, UNITE v10.0, DOI 10.15156/BIO/2959332; dynamic clustering, full-length ITS; Nilsson et al. (2019)). This release provides one representative sequence per SH code, each with a single unambiguous taxonomy string in its header. We used the same UNITE reference database for both the GlobalFungi taxonomy and the GenBank `vsearch` assignment described below, so both sources reference an identical SH definition.

1.2.3 GlobalFungi records

We downloaded the GlobalFungi v5 sample metadata file and the combined ITS1/ITS2 SH abundance matrix. The GlobalFungi dataset comes pre-populated with SH codes that are based on the same UNITE release described above, and a pipeline described at globalfungi.com. We retained Canadian samples by filtering for country = Canada; barcoding region in {ITS2, ITSboth}; non-manipulated conditions (`manipulated == "NO"`); and exclusion of substrate types irrelevant to EcM fungi (shoot, air, water, sediment). We pivoted the resulting wide abundance matrix to long format and removed zero-abundance records, then joined UNITE taxonomy by SH code.

1.2.4 GenBank records

We searched GenBank for fungal ITS sequences from Canada using the query `"Canada"[Country] AND "Fungi"[Organism] AND ("internal transcribed spacer"[All Fields] OR "ITS"[All Fields])`, with no date cutoff and no lifestyle keyword, so the search returns all matching fungal ITS records regardless of guild. We retrieved FASTA sequences and record metadata (coordinates, collection date, host, isolation source) in batches through the NCBI Entrez API (`rentrez`; Winter (2017)).

To make the GenBank component comparable to the GlobalFungi component (which we restricted to samples with a `barcoding_region` of ITS2 or ITSboth) we used ITSx v1.1.3 (Bengtsson-Palme et al. (2013); HMMER backend, Fungi profile set) to detect and extract the ITS1 and ITS2 sub-regions from each retrieved sequence. We excluded records with no detectable ITS2 or with an extracted ITS2 fragment shorter than 100 bp before SH assignment. We recorded whether each record also carried a genuine standalone ITS1 detection in an `its_region` column (`ITS2` or `ITSboth`), directly analogous to GlobalFungi's `barcoding_region`.

We matched every eligible record against UNITE on its extracted ITS2 fragment alone. The `its_region` flag is for provenance only and does not influence assignment. This differs from the GlobalFungi pipeline, which independently classified each ITS region. Our pipeline preserves a one-row-per-accession schema, so we do not achieve exact parity with GlobalFungi at the level of individual sub-region observations. To gauge whether discarding ITS1 measurably

costs resolution, our pipeline cross-tabulates each record's assignment outcome (below) against whether it carried a standalone ITS1 detection (`genbank_its1_resolution_crosstab.csv`).

We used `vsearch` v2.30.4 (Rognes et al. (2016); `usearch_global`) to compare sequences against the UNITE reference at 98.5% identity, matching GlobalFungi's documented SH-assignment criterion. We searched both strands exhaustively (`--strand both --maxaccepts 0 --maxrejects 0`) and retained all top-scoring hits per query (`--top_hits_only`) so that we could resolve ties. Where a query's top hits spanned more than one UNITE reference, we resolved the tie by taxonomic agreement. We compared named-species hits by species name, and "dark" hits (UNITE placeholders ending `_sp`) by SH code, following UNITE's own dark-taxon convention (Nilsson et al. (2019); Ryberg and Nilsson (2018)). When all tied hits agreed at that level, we assigned the record that SH (`taxonomic_resolution = "sh"`). When they disagreed but shared a genus, we retained the record at genus-only resolution (`taxonomic_resolution = "genus"`, with `sh_code` and `species` set to `NA` and the genus preserved), following the principle of climbing to the lowest unambiguous rank (Heeger et al. (2018)). We treated only records whose tied hits disagreed at genus level as genuinely ambiguous and excluded them (logged to `genbank_ambiguous_species_excluded.csv`). Downstream SH-richness counts therefore treat genus-resolved records as carrying no SH.

We obtained the taxonomy for SH-resolved records directly from the matched SH representative in the UNITE reference. We obtained the host taxon from the structured GenBank `host` qualifier where available, and used a fallback regex extraction of binomial names from the `isolation_source` field. We preserve the raw extracted string as `host_taxon_raw` (see below). We also assigned each GenBank record a `canada_basis` provenance category, recording the evidence for its inclusion (structured country field, coordinates, both, or search-only).

We addressed inconsistencies in raw host-plant names across the source datasets (the GenBank `host` qualifier and `isolation_source` field, and GlobalFungi's `dominant_plant_species` and `other_plant_species` fields) using a single shared normalization function, `canonicalize_host()` (`00_setup.R`), applied identically wherever we compare a raw host string against a species list (GenBank assembly here, and the Eltonian and Hutchinsonian analyses below). It reduces each raw string to a standardized "Genus species" (or bare genus) form, or to a missing value where the string cannot be resolved to a plausible Latin name. This step is necessary because host names in both sources are free text rather than a controlled vocabulary, so the same species may appear as a full binomial, a misspelling, an abbreviation, a common name, or with inconsistent capitalization or stray punctuation.

1.2.5 Combining sources and spatial validation

We combined the GlobalFungi and GenBank datasets by row-binding. For all records with parsed coordinates we applied a spatial-containment check against the GADM administrative boundary for Canada (see below), buffered by 2 km (applied in the Canada Albers Equal Area Conic projection, so the buffer distance is a true 2,000 m regardless of latitude), recording the

result in a logical `coord_in_canada` column (TRUE = validated inside the buffered boundary; FALSE = coordinates present but outside; NA = no coordinates); original coordinates are always preserved. The buffer absorbs GADM coastline-simplification and topology artefacts, which would otherwise exclude a handful of coastal and small-island records (e.g. Gulf of St. Lawrence) carrying `country == "Canada"` metadata. All spatial analyses filter on `coord_in_canada == TRUE`, while non-spatial enumerations (taxon richness, record counts) include all records, treating a GlobalFungi `country == Canada` label as authoritative even where coordinates fall marginally outside the buffered Canada polygon.

1.2.6 Assigning EcM guild status

We determined fungal EcM status at the genus level using FungalTraits (Põlme et al. (2020)): we retained records whose UNITE-assigned genus carries `primary_lifestyle = ectomycorrhizal` in FungalTraits (genus names lower-cased before matching), and carried two additional FungalTraits fields, EcM lineage and exploration type, into the final dataset. The resulting primary analysis dataset (`data_derived/emf_canada_em_only.csv`) has one row per (sample × SH code) detection.

1.2.7 Definition of a "site"

For analyses requiring georeferenced sampling units, we binned sample coordinates to three decimal places of latitude and longitude (~100 m at temperate latitudes), and refer to these bins as sampling "sites". This collapses potential GPS-precision noise across repeat readings at one general location. We used site-binning for the ecozone sampling-coverage counts; the Wallacean occurrence counts (below) are intentionally left unbinned so that each distinct coordinate contributes an occurrence.

1.3 Reference data preparation

1.3.1 Spatial layers

We obtained Canadian national, provincial, and territorial boundaries from GADM via `geodata`, a higher-resolution Canada boundary and major-lake polygons from Natural Earth via `rnaturalearth`, and Canadian ecoregions (National Ecological Framework for Canada) with their ecozone-name lookup. We reprojected layers to WGS84 for coordinate operations and to the Canada Albers Equal Area Conic projection for mapping, simplified them with `terra` (`simplifyGeom`) where appropriate, and wrote them to `data_derived/spatial/`.

1.3.2 FungalRoot host-determination rule

We determined EcM host status for plant species and genera from FungalRoot (Soudzilovskaia et al. (2020)) using two independent routes.

For the species-level route, we used Darwin Core Archive occurrence records from GBIF (DOI 10.15468/a7ujmj), joining the occurrence file to the "measurements" extension on the mycorrhiza-type field. We classified a species as EcM if FungalRoot carries at least one occurrence with an unambiguous EcM-positive label ("EcM", "EcM, no AM colonization", or "ErM,EcM"). Two further labels - "EcM, AM undetermined" and "EcM,AM" - do not on their own qualify a species as EcM-demonstrated, because they are too ambiguous to serve as positive evidence. However, records carrying these labels are not treated as evidence against EcM status either: a species with such a record is still classified as EcM-demonstrated if at least one other, unambiguous record for that species exists elsewhere in FungalRoot. We excluded `EcM, AM undetermined` because, in at least one case we identified (the fern *Dryopteris filix-mas*), it was contradicted by separate AM and non-mycorrhizal records for the same species, consistent with this label denoting an EcM report for which AM status was simply not surveyed, rather than a confirmed AM-negative result. We excluded "EcM,AM" (a dual-positive call, indicating both EcM and AM colonization were reported) as qualifying evidence at the species level because it was frequently the only EcM signal available for a species, with no unambiguous EcM record found anywhere else in that species' genus. Of an initial 698 candidate species that qualified under a broader rule, 118 species (spanning 20 genera, including *Acer*, *Allium*, *Fraxinus*, and *Juglans*) relied solely on this dual-positive label. Given the lack of any corroborating unambiguous evidence, we judged this label too weak on its own to support an EcM-demonstrated classification (but see below).

For the genus-level route, we used FungalRoot's own published genus-level recommendation, Supplementary Table S2 (Soudzilovskaia et al. (2020)), which assigns each genus a mycorrhizal-type call based on at least 67% consistency of species-level diagnosis. A genus qualifies as an EcM host genus if Table S2 assigns it "EcM" or "EcM-AM"; the latter captures genera with genuine, well-supported dual mycorrhizal associations (e.g., *Salix*, *Populus*, *Eucalyptus*) that the binary species-level occurrence rule cannot express. A handful of Table S2 rows carry non-standard free-text or apparent data-entry-error values (e.g., "species-specific: EcM or NM-AM"); these do not match either qualifying label, so we exclude them by default. The two routes are independent: a species/genus can qualify via its own direct occurrence record, via Table S2, or both, and a species need not have an occurrence record of its own to qualify if its genus is included via Table S2.

We acknowledge the species-level criterion is conservative. Some genera excluded under it, including *Acer*, *Fraxinus*, and *Juglans*, have published evidence of genuine EcM associations elsewhere in the literature (e.g., *Fraxinus* is listed among confirmed dual-mycorrhizal genera by Teste et al. (2020)), evidence that is not captured by any unambiguous occurrence record in the FungalRoot GBIF extract used here. The source Table S2 does not resolve this gap: it independently classifies *Acer*, *Fraxinus*, and *Juglans* as AM, so the genus-level route adds

no evidence for these three genera either. We chose not to supplement FungalRoot with case-by-case literature additions, to keep host determination reproducible from a single, clearly defined source; the cost of this choice is that our EcM host list under-counts true EcM hosts for a small number of genera whose FungalRoot evidence (occurrence- and genus-level alike) is exclusively ambiguous or absent.

1.3.3 Native Canadian EcM host species

We identified native EcM host plant species occurring in Canada by querying BIEN (Enquist et al. (2026); Maitner et al. (2018)) for the plant species with Canadian distributions, filtering to natives via the Native Species Resolver (NSR; Boyle et al. (2024)), one of BIEN's modular data services (Enquist et al. (2026)), which determines whether a species is native or introduced to the political division of the observation by comparing it against curated regional checklists. For Canadian determinations NSR draws on the Database of Vascular Plants of Canada (VASCAN; Brouillet (2025)), which NSR last refreshed on 16 September 2024, and the World Checklist of Vascular Plants, delivered via Plants of the World Online (WCVP v13; Govaerts (2024)), refreshed 21 May 2024. We queried NSR through its R package (Maitner and Boyle (2024)), and intersected the resulting native flora against the EcM host determination described above. We retain a native species if it qualifies under **either** route: its own species-level occurrence record satisfies the species rule, **or** its genus qualifies under FungalRoot's Table S2 genus rule. We combine the two routes with OR, as implemented in `06_host_species.R` (Step 4). The resulting list (`data_derived/ecm_native_canada_host_species.csv`) records which route(s) applied for each species (`evidence_source`: "occurrence", "table_s2", or "both") and is used in the Hutchinsonian and Eltonian analyses.

Combining the two routes with OR also mitigates a taxonomic-resolver limitation. Because BIEN's occurrence-name scrubbing, the Native Status Resolver, and the GBIF backbone underlying FungalRoot are independent taxonomic resolvers, a genuinely native, EcM-demonstrated species can fail the species-level intersection when no single name string clears both the native filter and the species-level demonstrated filter simultaneously. The green/mountain alder complex (*Alnus alnobetula* / *A. crispa* / *A. viridis*) is a case in point: the native filter returns "Native" only under *Alnus alnobetula* while FungalRoot's unambiguous species-level EcM evidence is filed under *Alnus crispa*, so the species-level route alone would not match. Here the genus route rescues the species: genus *Alnus* is called "EcM" in Table S2, so *Alnus alnobetula* is retained (via `evidence_source == "table_s2"`). A residual limitation remains only for species whose species-level name fails to match across resolvers **and** whose genus is absent from Table S2; we flag this as a known limitation of unresolved cross-database synonymy rather than applying a general synonym-harmonization step (which could shift counts for many species).

We assigned each host species a growth form (tree, shrub, or herbaceous), which we use for the tree / non-tree split reported in the main text and in Figure 4. Growth form comes from a four-stage chain, applied in order and recorded per species in the `growth_form_source` column of `data_derived/ecm_native_canada_host_species.csv`: (i) the modal value in the BIEN

trait database (Enquist et al. (2026); Maitner et al. (2018)); (ii) for species BIEN does not cover, GIFT trait 1.2.1 (Denelle and Weigelt (2024), queried at `agreement = 0.66`); (iii) for species neither database covers, the modal growth form of that genus's other species in this host list; and (iv) a curated manual override for any species still unresolved. Of the **272** host species, **215** growth forms come from BIEN, **15** from GIFT, **42** from the congener-modal step, and **0** from the manual override. We flag the congener-inferred group explicitly because those values are inferred from congeners rather than taken from a trait record for the species itself.

1.3.4 BIEN2 host ranges and richness rasters

We obtained MaxEnt-based modelled range polygons (BIEN2; Moulatlet et al. (2025)) for the native Canadian EcM host species from biendata.org, downloaded programmatically through the REST API, reprojected to WGS84, and clipped to the Canada boundary. We rasterized these polygons onto a 0.5° geographic grid to produce three layers: host species richness (count of host species with a range polygon per cell); host data richness (count of host species with at least one Canadian EcM molecular-detection record per cell); and host data proportion (the ratio of the two).

Not every host species contributes a range. Of the **272** native Canadian EcM host species, **227** have a BIEN2 modelled range that overlaps Canada and therefore appear in the richness rasters. The remaining **45** do not. Most of these have no BIEN2 model at all. A smaller number do have a modelled range, but one that does not reach the Canada boundary once clipped: BIEN2 ranges are statistical predictions built from occurrence records, and a species whose Canadian presence is a marginal, edge-of-range extension of a mostly non-Canadian distribution (for example *Cercis canadensis* and *Quercus muehlenbergii*, both native to Canada only at the northern limit of a range centred well to the south) can have too few Canadian records for the model to predict presence there, even though the species is genuinely native. Native status and a modelled range are therefore two different lines of evidence, and the second does not always follow from the first. The host-richness and host-proportion layers, and Figure 4, are therefore based on the **227** species with usable ranges rather than the full host list. This makes the mapped host coverage conservative: a cell may hold EcM host species whose ranges BIEN2 does not model.

1.4 Shortfall analyses

1.4.1 Linnean shortfall

We quantified the Canadian EcM fungal inventory at three taxonomic levels: unique UNITE SH codes, genera, and named species (excluding unresolved UNITE entries ending in `_sp`). SH codes are the primary richness unit because they provide a consistent operational taxonomic unit applicable to both sources, and because many EcM taxa are not yet formally named.

We compared observed Canadian EcM genera to the global EcM genus inventory in FungalTraits, and quantified the overlap between the GlobalFungi and GenBank components at the named-species SH level using set operations (shared, GlobalFungi-only, GenBank-only). We characterized the dataset "dark fraction" as the proportion of detected SH codes lacking a species-level UNITE name, computed for the combined dataset and for each source separately. This SH-level dark fraction is not directly comparable to the OTU-level dark-diversity estimates of van Galen et al. (2025), whose denominators include OTUs matched only to genus level.

We also computed the average per-sample proportion of species-unassigned ("dark") OTUs among van Galen et al.'s Canadian forest and tundra samples, using their published sample-specific estimates (Figshare DOI 10.6084/m9.figshare.28830371), a table giving, for each GlobalFungi v5 sample in their global compilation, the proportion of EcM OTUs assigned and unassigned to a named species, together with sample-level metadata including country and biome. We filtered this table to samples with country = Canada and biome in {forest, tundra} (the subset of their global sampling most comparable to our own) and took the unweighted mean across those samples. We report this in the main text alongside our own SH-level dark-fraction estimate as a rough, non-equivalent point of reference rather than a direct comparison: the two measures use different taxonomic units (OTU versus SH code) and different underlying clustering/assignment pipelines.

1.4.2 Wallacean shortfall

We characterized geographic sampling breadth for each EcM taxon at three levels (SH code, genus, named species). For each taxon we counted unique sampling occurrences as distinct 30 arc-second grid cells (~1 km^2), and produced frequency distributions of occurrences per taxon within Canada and globally. We report the proportion of taxa known from a single occurrence, and from fewer than 30 occurrences, both within Canada and globally, as straightforward measures of the shortfall. The within-Canada and global occurrence counts are drawn from the GlobalFungi dataset only, because it provides the structured sampling frame required for meaningful occupancy comparison; the global counts map each named Canadian species (or SH code) to all globally-known SH codes carrying that identity and match them against the full global GlobalFungi v5 SH abundance matrix under the same quality filters used for the Canadian data. As a complementary view relevant to species distribution modelling, we tabulated how many taxa meet common minimum-occurrence thresholds for modelling (30 occurrence records).

1.4.3 Prestonian shortfall

We obtained BioTIME (Dornelas et al. (2018)) directly from the BioTIME data portal (biotime.st-andrews.ac.uk) on 5 March 2026, downloading both the raw biodiversity time-series data (as an RDS file) and the accompanying study metadata (as a CSV) via the portal's direct-download endpoints. We searched this database for time-series records of EcM fungal

communities in Canada, filtering to studies with at least one Canadian location, at least two sampling time points, and an organism type consistent with fungi, and cross-referencing matching studies against the EcM dataset by taxonomy. We operationalize the Prestonian shortfall as the percentage of named EcM fungal species present in Canada that lack any abundance time-series data recorded anywhere on Earth.

1.4.4 Darwinian shortfall

We assessed genome-sequence availability for Canadian EcM genera by cross-referencing detected genera against the JGI MycoCosm fungal-genomics portal. We obtained MycoCosm's organism list on 5 March 2026 from its public "All Fungi" listing page (mycocosm.jgi.doe.gov/fungi/fungi.info.html, no login required), which lists every fungal genome in the portal; at the time of data collection the portal's automated bulk-export feature required a JGI login and was not functioning, so we retrieved the listing table by selecting and copying its contents directly from the page. Because the source page lists fungal genomes exclusively, we applied no additional kingdom- or phylum-level filtering. We operationalize the Darwinian shortfall as the percentage of named Canadian EcM fungal species absent from the database.

1.4.5 Raunkiæran shortfall

We quantified trait-data availability using FungalTraits (a genus-level compilation) for all EcM genera detected in Canada. We assessed the six FungalTraits columns carrying information meaningful for EcM fungi, in two classes: three morphological/structural traits derivable from taxonomic description (fruitbody type, hymenium type, mycorrhizal exploration type) and three ecological/interaction traits (secondary lifestyle, endophytic-interaction capability, host specificity). We excluded the remaining FungalTraits columns as inapplicable to EcM fungi by lifestyle (i.e., those describing decay and saprotrophic modes, lichen photobionts, plant pathogenicity, animal biotrophy, and aquatic habitat). The seventh potentially relevant column, growth form, is "filamentous mycelium" for all EcM genera, and we excluded it as uninformative. Coverage is the proportion of genus × trait cells with a documented value, where "documented" means non-missing and not the literal entry "unknown". Blanks in secondary lifestyle and endophytic-interaction capability are ambiguous between "not assessed" and "trait absent", so coverage for those two traits is a lower bound.

1.4.6 Hutchinsonian shortfall

We addressed environmental coverage at two levels.

For **climate-space coverage**, we characterized how completely Canada's climate space has been sampled. We defined climate space using WorldClim 2.1 bioclimatic variables BIO1

(mean annual temperature) and BIO12 (mean annual precipitation), aggregating the 30 arc-second rasters to ~10 arc-minutes and masking to the Canada boundary, then binned the two-dimensional temperature × precipitation space into a 50 × 50 grid of equal-width climate zones. For each zone we computed coverage as the proportion of its Canadian grid cells containing at least one EcM sample (each cell counted once), with a coverage of zero denoting a climate present in Canada but never sampled. We mapped coverage in both geographic and climate space, separately for GlobalFungi samples and for the combined GlobalFungi + GenBank dataset, against the available-climate frequency (the number of Canadian cells per zone). We adapted the code producing this figure from Hébert et al. (2026).

For **ecozone coverage**, we determined which of Canada's 15 terrestrial ecozones (National Ecological Framework for Canada; ecozones.ca) contain at least one EcM sampling location, and report the number and proportion of ecozones sampled at a range of site thresholds.

1.4.7 Eltonian shortfall

We first assessed the **availability of host information** in the raw sample metadata. For GlobalFungi, among Canadian samples carrying at least one EcM SH code we tallied how many had an entry in the `dominant_plant_species` field, and cross-tabulated the samples with host information by `sample_type` (soil vs. root) (`18_eltonian.R`, Step A2c). For GenBank, we binned each Canadian EcM record's `isolation_source` string into tissue categories using keyword matching (`18_eltonian.R`, Step A2d). Note that this availability diagnostic keys on the single `dominant_plant_species` field for GlobalFungi, whereas the fungus–host association analysis below draws host identities from **both** `dominant_plant_species` and `other_plant_species`; the two therefore measure subtly different quantities (presence of a primary host label vs. any usable host identity), so their sample counts need not coincide.

Tissue categories for GenBank records. The keyword rule assigns each `isolation_source` string, in priority order, to: *root / ectomycorrhizal tissue* (matches `root`, `mycorrhiz`, or the whole word `ecm`); *root and soil both mentioned* (matches a root keyword and also `soil`, `rhizosphere`, or `duff`); *soil / rhizosphere / duff*; *non-root tissue* (`basidiocarp`, `sporocarp`, `leaf`, `leaves`, `needle`, `bark`, `stem`, `thallus`, `macroalga`); *habitat description only* (text present but naming no material, for example a stand age or forest type); and *not recorded* (field empty). The "root and soil both mentioned" category is mechanical: every string it captures in this dataset describes root material collected where soil was also sampled (for example, "mycorrhizal root tips of *Pinus albicaulis* … and soil adjacent to *Pinus albicaulis*"), so these records are treated as root evidence. The distinction between *habitat description only* and *non-root tissue* also matters for interpretation: a habitat description tells us nothing about the material sampled, whereas a non-root tissue string tells us the material was not root. The two are therefore reported separately. Every category count is written to `eltonian_genbank_tissue_tally_canada.csv`, and the reported figures use this keyword scheme throughout.

We then focused on EcM fungus–host plant associations, in two scopes. For the **Canadian scope**, we extracted host information from the GenBank `host` qualifier (with fallback from `isolation_source`) and from GlobalFungi's `dominant_plant_species` and `other_plant_species` fields. Each database is restricted to root-derived evidence, but by a different mechanism, so the two rules are stated separately. For GlobalFungi, only samples with `sample_type == "root"` contribute host identities; the plant fields of soil and other samples record what grew at the site, not what the fungus was partnered with. For GenBank there is no controlled sample-type field, so we require the free-text `isolation_source` to match the root keyword rule described above, so that GenBank is held to the same evidential standard as GlobalFungi: without this restriction, records from soil, rhizosphere, habitat descriptions, and blank tissue fields would contribute host associations on the same footing as a root tip. Records excluded by the root-evidence rule remain in the taxonomic and geographic assessments; they are removed only from the association analyses.

Multi-species host labels. Some metadata fields name several plants for one sample, for example "*Picea mariana*, *Picea glauca*, *Pinus banksiana*". The shared `canonicalize_host()` helper reduces a string to its first two tokens, so a multi-plant label has to be split into separate candidates before that helper is applied; without this step every plant after the first would be silently discarded. We split each label on commas and semicolons and treat every named plant as a separate candidate host. Because a fungus detected in a sample that named eight plants is weaker evidence than one detected where a single plant was named, each record carries a confidence flag: *unambiguous* if the source record named exactly one candidate plant, *ambiguous* otherwise. For GlobalFungi the count is taken across both plant fields together, so a sample whose `dominant_plant_species` names one species but whose `other_plant_species` names seven more is correctly flagged ambiguous. A host × fungus pair is called unambiguous if any one of its supporting records was. All pairs are retained in the reported analysis; the confidence split is written to `eltonian_pair_confidence.csv`, and `eltonian_sensitivity_confidence.csv` reports the headline Canadian statistics recomputed from unambiguous evidence alone. We canonicalized all host strings with the shared helper and matched them against the native Canadian EcM host species list, which serves as the denominator of potential interactions. We built fungus × host interaction matrices at the SH-code, named-species, and genus levels, and computed the proportion of Canadian EcM host species with at least one documented fungal association, the fill rate of the full host × named-fungal-species matrix, and the proportion of observed pairs supported by only a single record. For the **global scope**, we queried the full GlobalFungi global metadata for root samples matching Canadian host species, and queried GenBank globally for each Canadian EcM genus (retrieving all matching records), to establish which Canadian hosts and fungi have any documented association anywhere. To contextualize the sampling gap relative to host diversity, we produced a bivariate map classifying each 0.5° grid cell by EcM host species richness versus the proportion of those host species with EcM molecular-detection data in Canada, shown separately for tree and non-tree hosts (this bivariate map is produced in `18_eltonian.R`).

Sensitivity check: is the full matrix an appropriate denominator? Reporting the complete host × fungal-species grid treats every pairing as a candidate interaction. We describe

this grid as a *metaweb* (the set of pairings conceivable given the species present), recognizing that not every cell is geographically feasible. The script `21_geographic_possibility.R` explores how alternative definitions of the matrix affect the percentage of cells filled. A host species counts as geographically possible for a fungal species if at least one of that fungus's georeferenced Canadian occurrence records falls inside the host's modelled BIEN range. The calculation is repeated at six increasingly permissive spatial grains: point inside the range polygon; point within 25 km or 100 km of it (`sf::st_is_within_distance()` in the Canada Albers equal-area projection); point in a 0.5° grid cell the host occupies; and point in an ecoregion or an ecozone the host's range intersects. Results are written to `data_derived/geo_possibility/`.

We report this filter as a sensitivity check rather than adopt it as our primary denominator. The result depends heavily on the grain chosen, from **28.2**% of cells surviving the strictest definition to **50.0**% surviving the most permissive. It is also more a measure of survey effort than of true geographic incompatibility: a fungus known from only a few locations will always have a short possible-host list, whatever its real range. Much of what the filter removes, moreover, is missing data rather than geography (**243** named fungal species have no georeferenced Canadian record and **45** host species have no BIEN range, together accounting for **43.1**% of the cells removed at the finest grain), and the range maps themselves carry error: **2.6**% of the pairs we have actually observed and can test still come out "impossible" under this filter (concentrated in *Pseudotsuga menziesii* and *Quercus garryana*, both host species with restricted modelled ranges). Because shrinking the denominator while the numerator stays fixed would make the shortfall look artificially less severe, we retain the full, unrestricted matrix as our primary result.

1.4.8 Extra analyses

We compiled physical EcM specimen records from GBIF as a vouchered complement to the sequence data: an occurrence download for Kingdom Fungi in Canada (basis of record `PRESERVED_SPECIMEN` or `LIVING_SPECIMEN`; records with geospatial issues excluded), filtered to species-rank records in EcM genera. For the primary map (Figure S5), the "with-sequence" genera are those observed in our own Canada-specific dataset (itself restricted to FungalTraits EcM genera), while Table S4's "without-sequence" genera are drawn directly from the FungalTraits EcM genus list (`09_linnean.R`). The download (retrieved 6 March 2026; 673,532 occurrence records across 294 datasets before filtering) is archived at https://doi.org/10.15468/dl.92rns5 (download key `0017649-260226173443078`; predicates country = CA, `HAS_GEOSPATIAL_ISSUE` = FALSE, taxonKey = 5, `BASIS_OF_RECORD` in {`PRESERVED_SPECIMEN`, `LIVING_SPECIMEN`}, format `SIMPLE_CSV`; licence CC BY-NC 4.0).

1.5 Figure preparation

Every figure in the manuscript and in these supplemental materials is generated directly by the analysis scripts, with one exception described below. Each analysis script writes its figure

in two versions: a white-background version, which is the version reproduced in the manuscript and supplements, and an otherwise identical version on a light grey background (`#F2F2F2`).

We used the grey-background versions to assemble Figure 5, which is a designed conceptual composite rather than a computed figure. We laid out its five source panels manually in Microsoft PowerPoint and added the annotations that appear in it: the quadrant labels and the ellipses that link sampling strategies to the shortfalls they would jointly address. Figure 5 is therefore the only figure that cannot be regenerated by running the pipeline. Its source panels can be, and are listed below.

Source panels used to assemble Figure 5, and the analysis script that produces each.

Source panel (grey-background version)	Produced by
Figure-01_sampling_map_grey.png	19_sampling_maps.R
Figure-03_climate_gap_grey.png	17_hutchinsonian.R
Figure-04_host_bivariate_map_grey.png	18_eltonian.R
Figure-S1_dark_diversity_grey.png	10_dark_diversity.R
Figure-S4_ecozone_sampling_map_grey.png	17_hutchinsonian.R

`19_sampling_maps.R` also produces a grey-background version of Figure S5 (`Figure-S5_gbif_specimens_grey.png`), following the same convention as the panels above, but it is not used in Figure 5.

2 Supplemental Tables

Table S1. Unique EcM fungal sampling locations per Canadian ecozone (National Ecological Framework), for the combined GlobalFungi + GenBank dataset (n = 1,666 locations across 15 ecozones) and for GlobalFungi only (n = 486 locations), with ecozone area and combined sampling density. Points are assigned to ecozones by a within-polygon join against the detailed (unclipped) ecoregion layer, with points falling just offshore of the ecoregion coastline snapped to the nearest ecozone within 5 km; ecozone areas use the same polygon basis, so counts and areas are consistent. Ordered by decreasing combined density.

Ecozone	GlobalFungi (loc.)	Total (loc.)	Area (km²)	Density (loc. / 10,000 km²)
Mixedwood Plains	58	406	169258	23.99
Pacific Maritime	2	213	251087	8.48
Atlantic Maritime	104	163	282969	5.76
Boreal Shield	222	678	2079895	3.26
Montane Cordillera	17	62	489977	1.27
Taiga Plains	46	47	657947	0.71
Boreal Plains	20	38	740741	0.51
Boreal Cordillera	6	16	470758	0.34
Taiga Shield	9	24	1453302	0.17
Arctic Cordillera	1	2	340729	0.06
Northern Arctic	0	13	2838182	0.05
Prairies	1	2	466981	0.04
Southern Arctic	0	2	1683284	0.01
Hudson Plains	0	0	448432	0.00
Taiga Cordillera	0	0	266972	0.00

Table S2. The 28 FungalRoot/BIEN native Canadian EcM host species with at least one recorded, species-level host taxon in Canadian GlobalFungi or GenBank EcM records (the intersection between the native host list and the distinct species-level host taxa recorded across both sources). The second column shows whether that host name is carried by at least one root-derived record (a GlobalFungi root sample, or a GenBank record whose isolation source names root or ectomycorrhizal material). Only the 20 species marked "Yes" contribute to the fungus–host association results; the remaining 8 are named only by records whose tissue provenance is soil, non-root, or unstated.

Host species	Root-derived evidence
Abies balsamea	Yes
Alnus alnobetula	Yes
Alnus incana	No
Betula nana	Yes
Betula papyrifera	Yes
Dryas integrifolia	Yes
Gaultheria shallon	No
Monotropa uniflora	No
Picea engelmannii	Yes
Picea glauca	Yes
Picea mariana	Yes
Picea rubens	Yes
Pinus albicaulis	Yes
Pinus banksiana	Yes
Pinus contorta	Yes
Pinus ponderosa	Yes
Pinus rigida	No
Pinus strobus	Yes
Populus balsamifera	Yes
Populus tremuloides	Yes
Populus trichocarpa	No
Pseudotsuga menziesii	Yes
Pterospora andromedea	No
Quercus garryana	Yes
Quercus rubra	No
Salix arctica	Yes
Salix reticulata	No
Tsuga heterophylla	Yes

Table S3. The 31 species-level host names recorded in the Canadian GlobalFungi or GenBank metadata that did not match the FungalRoot/BIEN native EcM host species list (Table S2). These include ornamental or non-native species, misspellings, and other problematic or truncated entries, and are excluded from the host-interaction analyses. Both this table and Table S2 are built from every host name the metadata carries, regardless of tissue provenance, because their purpose is to document the quality of the recorded host names rather than the evidence available for interactions.

Host name (metadata)
Abies balsamifera
Alnus crispa
Aralia nudicaulis
Ascophyllum nodosum
Betula in
Betula payrifera
Chamaedaphne calyculata
Cornus canadensis
Corylus avellana
Deschampsia cespitosa
Douglas fir
Fagus or
Fagus sylvatica
Gammarus tigrinus
Hyalella azteca
Kalmia angustifolia
Linnaea borealis
Myrica gale
Not available
Oak beech
Picea abies
Pinus mugo
Pinus radiata
Populus alba
Populus and
Populus in
Quercus robur
Rosa acicularis
Ulva intestinalis
Viburnum edule
Zostera marina

Table S4. The 32 EcM fungal genera (FungalTraits) represented in the Canadian GBIF physical-specimen records (Figure S5) that have no sequence data in Canada (not detected in the combined GlobalFungi + GenBank dataset), and so could be targeted for future molecular sampling. 'GBIF records' and 'GBIF species' are the count of physical-specimen occurrence records and distinct species names per genus in the GBIF download.

Genus	GBIF records	GBIF species
Boletinus	276	5
Turbinellus	190	2
Imleria	125	2
Gomphus	115	1
Albatrellopsis	109	3
Xanthoconium	84	2
Wynnella	41	2
Hortiboletus	33	2
Boletellus	27	1
Laeticutis	21	1
Gyrodon	19	1
Aleurina	18	3
Baorangia	14	1
Tremellodendron	14	3
Phallogaster	12	1
Cyanoboletus	11	3
Neoboletus	11	3
Underwoodia	10	1
Gastroboletus	6	2
Paragyrodon	6	1
Bankera	5	2
Boletochaete	5	1
Pseudocraterellus	5	2
Riessia	4	1
Porpoloma	3	1
Hydnangium	2	1
Plicariella	2	2
Zelleromyces	2	2
Brauniellula	1	1
Marcelleina	1	1
Mycolevis	1	1
Protoglossum	1	1

3 Supplemental Figures

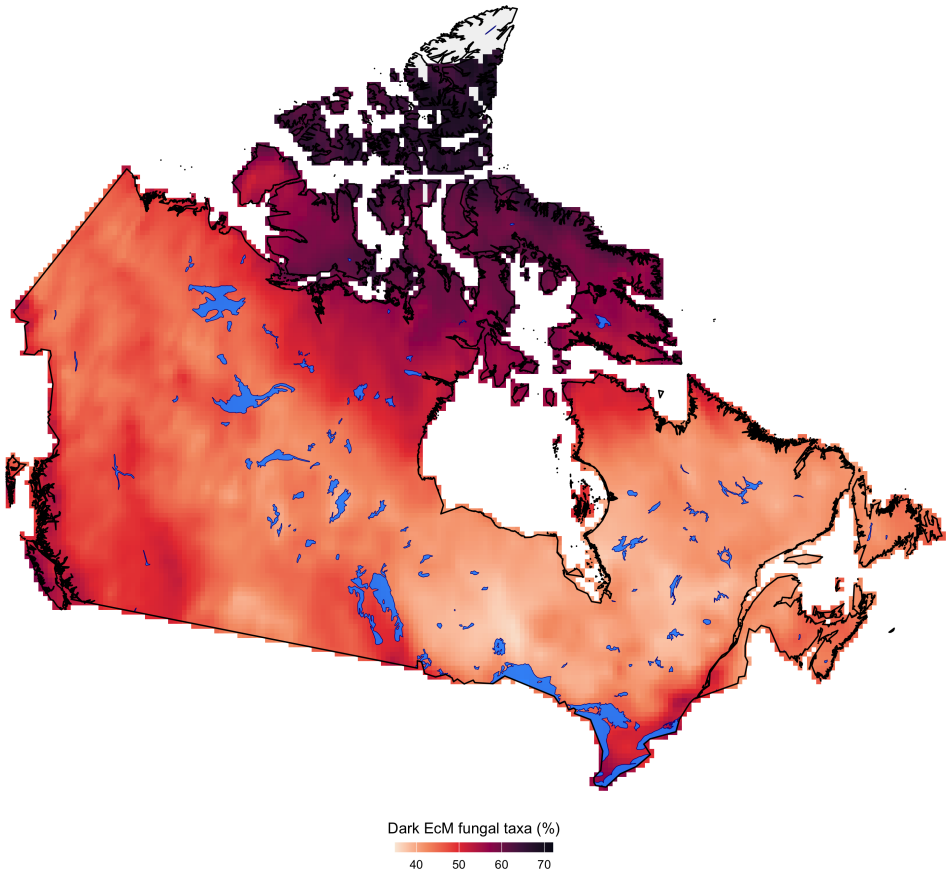


Figure S1. Modelled percentage of EcM OTUs that are undescribed across Canada, clipped from the global geospatial dataset of van Galen et al. (2025).

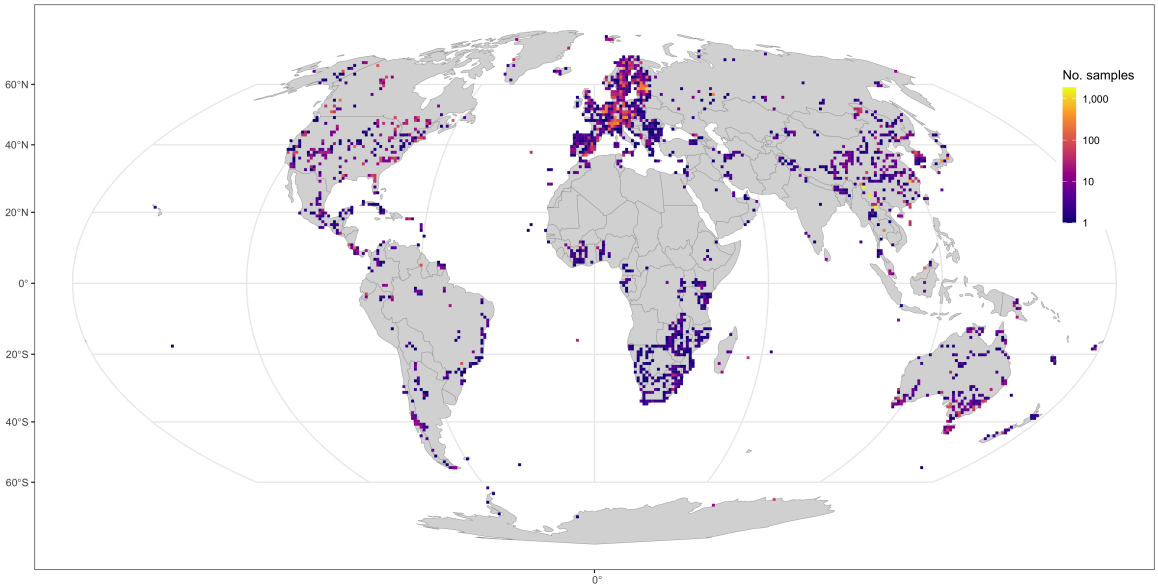


Figure S2. Sampling density of GlobalFungi samples in which at least one EcM fungal taxon was detected (n = 34,925; in Canada, n = 1,469 samples). Each coloured pixel represents a single 100 × 100 km region.

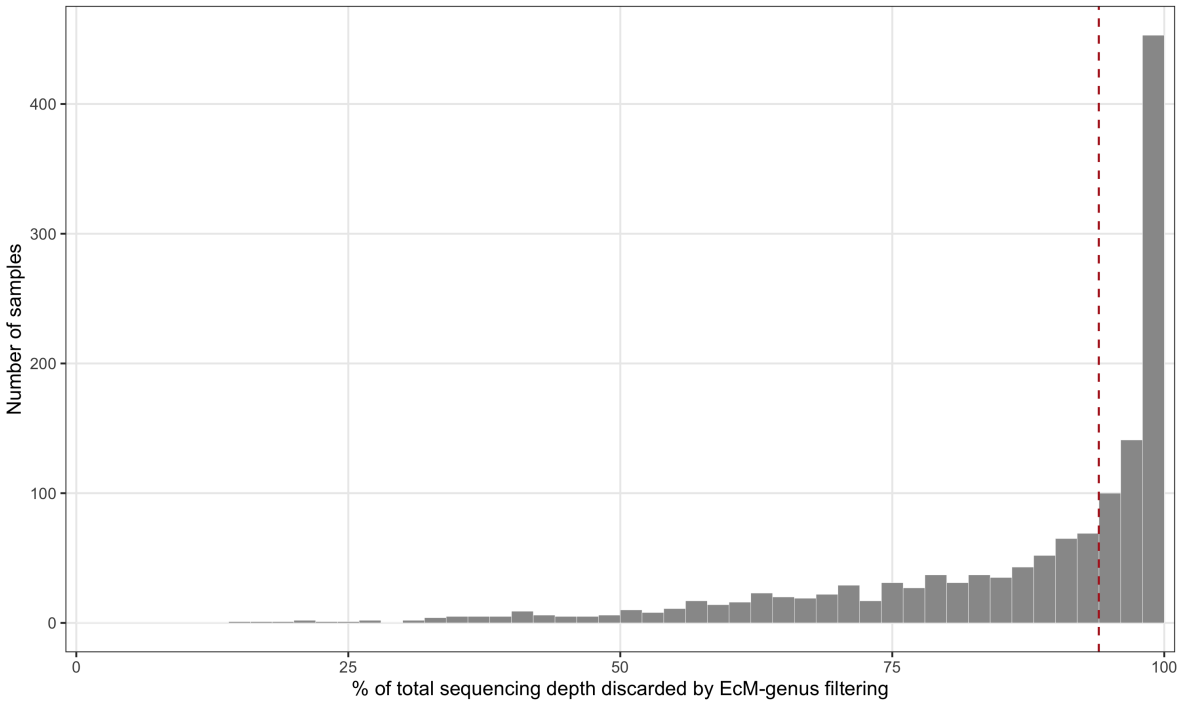


Figure S3. Distribution of the percentage of each Canadian GlobalFungi soil sample's total sequencing depth discarded by EcM-genus filtering, for every sample with at least one EcM-classified read (n = 1,388 samples). Q1–median–Q3: 79.2–94.0–99.0%. The vertical dashed line marks the median (94%).

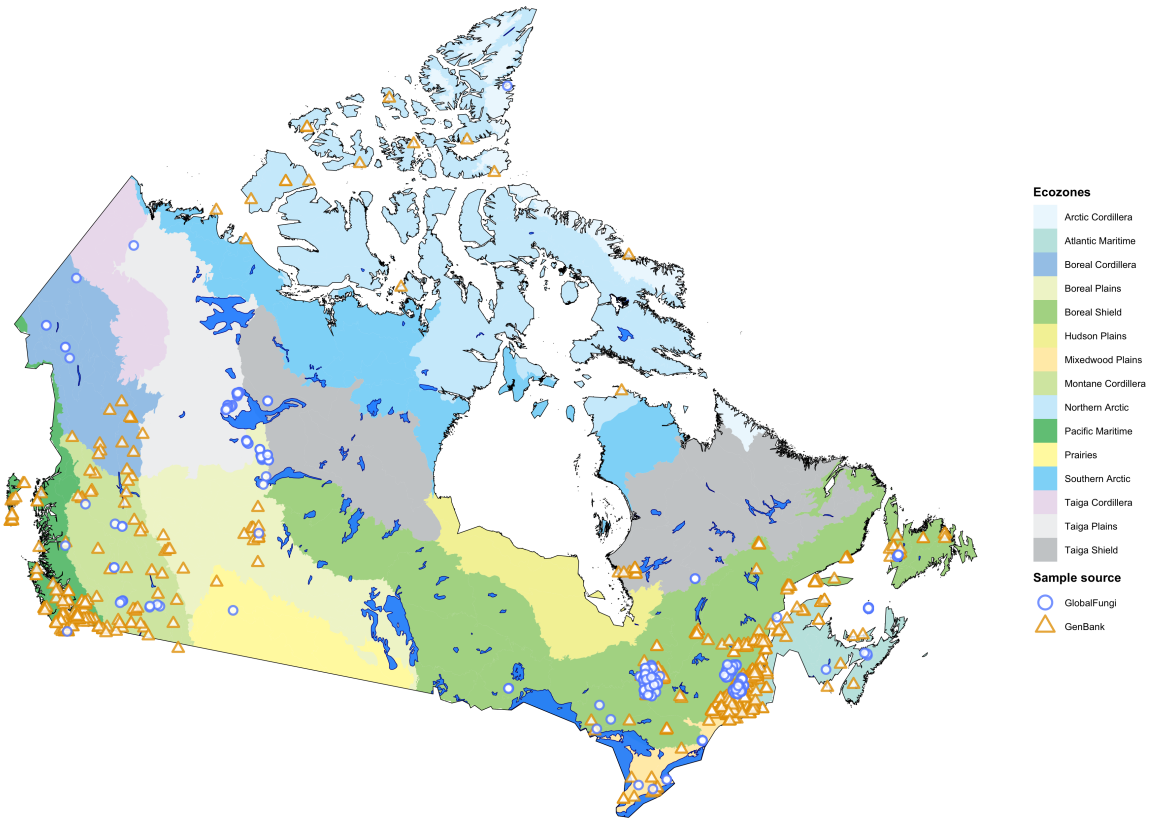


Figure S4. EcM fungal sampling locations (GlobalFungi and GenBank) projected in Canada Albers Equal Area and overlaid on Canadian ecozone polygons (National Ecological Framework for Canada). Sampling unit: one distinct raw lat/lon coordinate per record (GlobalFungi: n = 486; GenBank: n = 1,180; combined: n = 1,666 unique locations). Blue circles: GlobalFungi; orange triangles: GenBank.

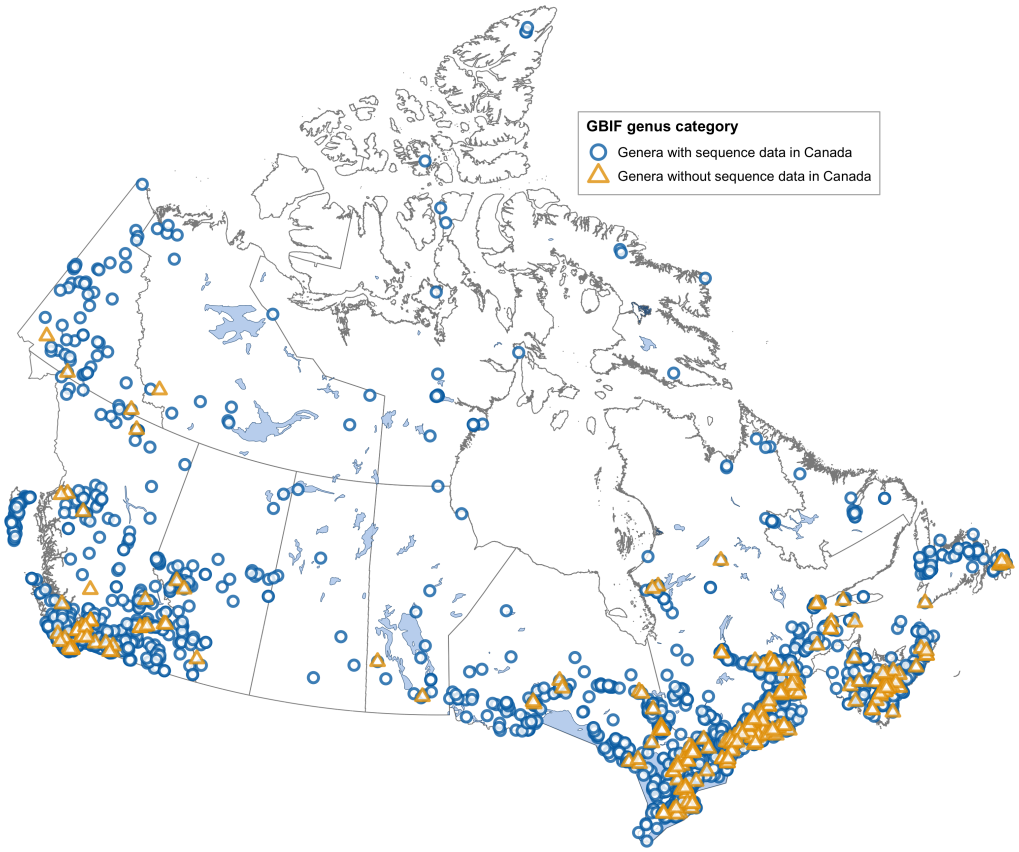


Figure S5. Occurrence records associated with physical EcM fungal specimen records in Canada, retrieved from GBIF on 6 March 2026 (`PRESERVED_SPECIMEN` and `LIVING_SPECIMEN`; unique occurrence records). Blue circles: occurrence records of EcM fungal genera (FungalTraits) with sequence data in Canada (combined GlobalFungi + GenBank; n = 4,131 plotted records). Orange triangles: occurrence records of EcM fungal genera without any sequence data in Canada (n = 299 plotted records). Plotted records are the georeferenced, distinct-coordinate occurrences within the Canada boundary, far fewer than the total specimen records because most GBIF records lack coordinates or share a coordinate. Source: GBIF Occurrence Download https://doi.org/10.15468/dl.92rns5.

References

Bengtsson-Palme J et al. 2013. Improved software detection and extraction of ITS1 and ITS 2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods in Ecology and Evolution, 4(10): 914–919. https://doi.org/10.1111/2041-210X.12073.

Boyle BL, Maitner B, Barbosa GC, Rethvick SYB, and Enquist BJ. 2024. Native Species Resolver. Available from https://nsr.biendata.org/.

Brouillet L. 2025. Database of Vascular Plants of Canada (VASCAN). Available from https://doi.org/10.5886/zw3aqw.

Chamberlain S et al. 2024. Rgbif: Interface to the Global Biodiversity Information Facility API.

Denelle P, and Weigelt P. 2024. GIFT: Access to the Global Inventory of Floras and Traits (GIFT).

Dornelas M, Antão LH, Moyes F, Bates AE, Magurran AE, et al. 2018. BioTIME: A database of biodiversity time series for the Anthropocene. Global Ecology and Biogeography, 27(7): 760–786. https://doi.org/10.1111/geb.12729.

Enquist BJ et al. 2026. BIEN: A biodiversity informatics ecosystem advancing open and reproducible workflows for plant observation, plot and trait data. Methods in Ecology and Evolution, 17(5): 1556–1584. https://doi.org/10.1111/2041-210X.70274.

Govaerts R. 2024. WCVP: World Checklist of Vascular Plants, Version 13. Available from http://sftp.kew.org/pub/data-repositories/WCVP/.

Hébert K et al. 2026. Blitz the Gap: A nation-wide effort to guide citizen science toward the needs of biodiversity science.

Heeger F et al. 2018. Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. Molecular Ecology Resources, 18(6): 1500–1514. https://doi.org/10.1111/1755-0998.12937.

Hijmans RJ. 2025. Terra: Spatial data analysis.

Hijmans RJ. 2026. Geodata: Access geographic data.

Maitner B, and Boyle B. 2024. NSR: 'Native Status Resolver'.

Maitner BS et al. 2018. The BIEN R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. Methods in Ecology and Evolution, 9(2): 373–379. https://doi.org/10.1111/2041-210X.12861.

Massicotte P, and South A. 2026. Rnaturalearth: World map data from Natural Earth.

Moulatlet GM et al. 2025. General laws of biodiversity: Climatic niches predict plant range size and ecological dominance globally. Proceedings of the National Academy of Sciences, 122(46): e2517585122. https://doi.org/10.1073/pnas.2517585122.

Nilsson RH et al. 2019. The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research, 47(D259–D264): 259–264. https://doi.org/10.1093/nar/gky1022.

Pebesma E. 2018. Simple Features for R: Standardized support for spatial vector data. The R Journal, 10(1): 439–446. https://doi.org/10.32614/RJ-2018-009.

Pedersen TL. 2024. Patchwork: The composer of plots.

Põlme S et al. 2020. FungalTraits: A user-friendly traits database of fungi and fungus-like stramenopiles. Fungal Diversity, 105(1): 1–16. https://doi.org/10.1007/s13225-020-00466-2.

R Core Team. 2026. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rognes T, Flouri T, Nichols B, Quince C, and Mahé F. 2016. VSEARCH: A versatile open source tool for metagenomics. PeerJ, 4: e2584. https://doi.org/10.7717/peerj.2584.

Ryberg M, and Nilsson RH. 2018. New light on names and naming of dark taxa. MycoKeys, 30: 31–39. https://doi.org/10.3897/mycokeys.30.24376.

Soudzilovskaia NA et al. 2020. FungalRoot: Global online database of plant mycorrhizal associations. New Phytologist, 227(3): 955–966. https://doi.org/10.1111/nph.16569.

Teste FP, Jones MD, and Dickie IA. 2020. Dual-mycorrhizal plants: Their ecology and relevance. New Phytologist, 225(5): 1835–1851. https://doi.org/10.1111/nph.16190.

van Galen LG et al. 2025. The biogeography and conservation of Earth's "dark" ectomycorrhizal fungi. Current Biology, 35(11): R563–R574. https://doi.org/10.1016/j.cub.2025.03.079.

Větrovský T et al. 2020. GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. Scientific Data, 7(1): 228. https://doi.org/10.1038/s41597-020-0567-7.

Wickham H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

Wickham H. 2026. Httr: Tools for working with URLs and HTTP.

Winter DJ. 2017. Rentrez: An R package for the NCBI eUtils API. The R Journal, 9(2): 520–526. https://doi.org/10.32614/RJ-2017-058.

Supplemental materials (SM2) for: An assessment of biodiversity data shortfalls for ectomycorrhizal fungi in Canada

Jason Pither, with assistance from Claude (Sonnet 5 and Opus 4.8)

Table of contents

1 Decisions log

Purpose
This log inventories the decisions we made in assembling and analyzing the dataset (Tables A1–A6), and, separately, the upstream decisions outside our control that are inherent to the external datasets we rely on (Tables B1 - B9).

How to read the tables
Each row describes one decision. The *Choice* column records what we did; the *Rationale* column records the stated

reason where one exists in the code or Supplemental Materials. Parameter values in `code font` are taken directly from the pipeline scripts; the parenthetical script reference (e.g. `02_globalfungi.R`) points to the source.

1.1 PART A — Decisions made by us

1.1.1 Table A1. GlobalFungi record assembly (`02_globalfungi.R`)

#	Decision	Choice	Rationale
A1	Barcoding-region filter	Retain `barcoding_region %in% {ITS2, ITSboth}`	Restricts to samples carrying an ITS2 read, the region used for SH assignment; sets up parity with the GenBank ITS2 restriction (Detailed methods SM1).
A2	Manipulation filter	Retain `manipulated == "NO"`	Exclude experimentally manipulated samples so records reflect natural communities (Detailed methods SM1).
A3	Substrate exclusion	Drop substrate in `{shoot, air, water, sediment}`	These substrates are irrelevant to (soil/root-associated) EcM fungi (Detailed methods SM1).

#	Decision	Choice	Rationale
A4	Species-level GlobalFungi matrix	**Not used**; work from the SH-level matrix only, decoded through the pinned UNITE build	Keeps a single, self-consistent SH definition across the project (`data_raw/DATA-DICTIONARY.md`).

1.1.2 Table A2. GenBank record assembly (`03_genbank.R`)

#	Decision	Choice	Rationale
B1	Search query	`"Canada"[Country] AND "Fungi"[Organism] AND ("internal transcribed spacer"[All Fields] OR "ITS"[All Fields])`; **no** date cutoff, **no** lifestyle keyword	Returns all matching fungal ITS records regardless of guild; we apply EcM filtering later and consistently (Detailed methods SM1).
B2	Sub-region extraction	ITSx (HMMER backend, Fungi profile set) to detect/extract ITS1 and ITS2	GenBank-side analogue of GlobalFungi's ITS2/ITSboth restriction, for cross-source comparability (Detailed methods SM1).
B3	ITS2 minimum length	Exclude ITS2 fragment `< 100 bp`	Floor to exclude spurious very short calls that trivially tie with many unrelated UNITE references (`03_genbank.R`, Detailed methods SM1).

#	Decision	Choice	Rationale
B4	SH matching input	Match on the **extracted ITS2 fragment only**; ITS1 not independently classified	Deliberate simplification preserving a one-row-per-accession schema; we disclose the consequence (loss of exact sub-region parity with GlobalFungi) and diagnose it via `genbank_its1_resolution_crosstab.csv` (Detailed methods SM1).
B5	`vsearch` identity threshold	`usearch_global` at **98.5% identity**	Matches GlobalFungi's documented SH-assignment criterion (Detailed methods SM1).
B6	Tie resolution — named vs dark	Named-species hits compared by species name; dark hits (`_sp` placeholders) compared by SH code	Follows UNITE's own dark-taxon convention (Nilsson et al. (2019); Ryberg and Nilsson (2018)) (Detailed methods SM1).
B7	Tie resolution — fallback rank	Agree at SH → assign SH; disagree but share genus → keep at **genus-only** (`sh_code`/`species` = NA); disagree at genus → **exclude** as ambiguous	"Climb to the lowest unambiguous rank" principle (Heeger et al. (2018)); excluded records logged to `genbank_ambiguous_species_excluded.csv` (Detailed methods SM1).

B8	Host taxon source	Structured `host` qualifier, with fallback regex extraction of binomials from `isolation_source`	`host` is the most reliable field; `isolation_source` recovers additional hosts (Detailed methods SM1).

1.1.3 Table A3. Host-name normalization — shared across sources (`canonicalize_host()`, `00_setup.R`)

#	Decision	Choice	Rationale
C1	Single shared normalizer	One `canonicalize_host()` applied identically wherever we match a raw host string to a species list (GenBank host field; Eltonian; Hutchinsonian)	Raw host names are free text (binomials, misspellings, abbreviations, common names, inconsistent case/punctuation); one shared function avoids ad hoc per-analysis cleaning and silent match failures (Detailed methods SM1).

#	Decision	Choice	Rationale
C2	Normalization steps	Strip surrounding punctuation → collapse whitespace → keep first two tokens → strip per-token punctuation → apply curated typo lookup → case-normalize (Titlecase genus, lowercase epithet) → reject non-Latin-name patterns → NA	Reduce each raw string to a standardized "Genus species"/"Genus" form, or NA if unresolvable (`00_setup.R`). Because the "keep first two tokens" step discards everything after the first species name, host labels that list several plants are split before this function is applied (decision F22b); the function itself is unchanged.
C3	Typo-correction lookup	Curated `typo_fixes` (currently `Abies balamifera → Abies balsamea`), extended as new misspellings surface	Manual, auditable corrections rather than fuzzy auto-correction (`00_setup.R`).

1.1.4 Table A4. Source combination, spatial validation, EcM guild, and site/occurrence grain conventions (`04_combine_ecm_dataset.R`, `00_setup.R`)

#	Decision	Choice	Rationale
D1	Coordinate validation	Spatial-containment check vs GADM Canada boundary, **buffered by 2 km** (in the Canada Albers Equal Area Conic projection) → logical `coord_in_canada` (TRUE / FALSE / NA); original coordinates always preserved	Distinguishes validated-inside from present-but-outside vs no-coordinates, without discarding data. The 2 km buffer absorbs GADM coastline-simplification and topology artefacts that would otherwise exclude a handful of genuinely Canadian coastal and island GlobalFungi records (Detailed methods SM1).
D2	Spatial vs non-spatial use	Spatial analyses filter `coord_in_canada == TRUE`; non-spatial enumerations (richness, counts) include **all** records	For counts, we treat a GlobalFungi `country == Canada` label as authoritative even where coordinates fall marginally outside the buffered polygon (Detailed methods SM1).
D3	EcM guild assignment	Genus-level: retain records whose UNITE-assigned genus is `primary_lifestyle == "ectomycorrhizal"` in FungalTraits (genus lower-cased before matching)	Genus-level EcM status from a single authoritative trait source (Detailed methods SM1).

#	Decision	Choice	Rationale
D4	FungalTraits version	**Pinned** to v1.2 (`FungalTraits_1-2.csv`); one file used by all scripts via `paths$fungaltraits`	Held fixed rather than tracked to a live source, so every script shares an identical EcM genus definition; provenance and SHA256 recorded in `fungaltraits_version.txt` (mirrors the UNITE build-pinning rationale given in Supplemental Materials SM1).
D5	"Site" definition	Bin lat/lon to **3 decimal places** (~100 m); `add_site_id()`	Collapses GPS-precision noise across repeat readings at one location without merging distinct sites (`00_setup.R`, Detailed methods SM1).
D6	Occurrence grain (Wallacean)	Distinct **30 arc-second** (~1 km²) grid cells; `snap_30arcsec()`	Aggregates occurrences to ~1 km² cells matching common SDM practice (`00_setup.R`, Detailed methods SM1).

#	Decision	Choice	Rationale
D7	Where we apply 3-decimal site-binning	We use the `add_site_id()` 3-decimal site-binning (D5) for the ecozone sampling-coverage counts, but do **not** apply it to the Wallacean per-taxon location table, whose within-Canada occurrence counts use raw distinct coordinates; we use the separate **30 arc-second** occupancy grid (D6) for the worldwide occupancy comparison and Figure 2	Distinct raw coordinates preserve per-record occurrences for the Canada location counts, while the coarser 30 arc-second grid gives a like-for-like occupancy comparison against the worldwide GlobalFungi dataset (`11_wallacean.R`, Detailed methods SM1).

1.1.5 Table A5. Reference / host-data preparation (`05_prepare_fungalroot.R`, `06_host_species.R`, `07_bien2_ranges.R`, `08_host_rasters.R`)

#	Decision	Choice	Rationale
E1	Host determination — two routes	Species-level (FungalRoot occurrence records) **OR** genus-level (FungalRoot Table S2); combined with OR	Two independent routes; a species/genus can qualify via either or both (Detailed methods SM1).

#	Decision	Choice	Rationale
E2	Qualifying species-level labels	Count as EcM-demonstrated only unambiguous EcM-positive labels: `EcM`, `EcM, no AM colonization`, `ErM,EcM`	Restrict to unambiguous positive evidence (Detailed methods SM1).
E3	Excluded label — `EcM, AM undetermined`	Excluded as qualifying evidence (but not disqualifying)	Denotes an EcM report with AM status unsurveyed, not a confirmed AM-negative; contradicted in a checked case (*Dryopteris filix-mas*) (Detailed methods SM1).
E4	Excluded label — `EcM,AM` (dual-positive)	Excluded as qualifying evidence	On inspection it was the sole EcM evidence for 118 of 698 candidate species across 20 genera (e.g. *Acer*, *Allium*, *Fraxinus*, *Juglans*) with no unambiguous record elsewhere (Detailed methods SM1).
E5	Genus-level qualifying labels	Table S2 assignment of `EcM` or `EcM-AM` qualifies	`EcM-AM` captures genuine dual-mycorrhizal genera (*Salix*, *Populus*, *Eucalyptus*) the binary species rule cannot express; at least 67% species-consistency basis is FungalRoot's own (Detailed methods SM1).

#	Decision	Choice	Rationale
E6	Non-standard Table S2 rows	Excluded by default (e.g. "species-specific: EcM or NM-AM")	Do not match either qualifying label (Detailed methods SM1).
E7	No literature supplementation	Do **not** add case-by-case literature EcM records (e.g. *Acer*, *Fraxinus*, *Juglans*)	Keep host determination reproducible from a single defined source; disclosed cost is under-counting a few genera (Detailed methods SM1).
E8	Native host list construction	BIEN species with Canadian distributions → filter to native via Native Status Resolver → retain species qualifying under **either** route in E1 (species-level occurrence list **OR** Table S2 genus list), combined with OR	Applies the two-route host determination (E1) to the native flora; no growth-form restriction. `evidence_source` records which route(s) applied (`06_host_species.R`, Step 4; Detailed methods SM1).

#	Decision	Choice	Rationale
E9	Growth-form assignment	Four-stage chain, applied in order: BIEN trait database → GIFT trait 1.2.1 (`agreement = 0.66`) → modal growth form of congeners already in the host list → curated manual override. The stage that supplied each value is recorded in `growth_form_source`	Growth form drives the tree / non-tree split reported in the main text and Figure 4, so it must be resolved for every host species; BIEN alone covers 215 of 272. Recording the source keeps the 42 congener-inferred values distinguishable from values taken from a trait record for the species itself (`06_host_species.R`, Steps 5–7c).

#	Decision	Choice	Rationale
E10	Unresolved synonymy	Accept drop-outs from independent taxonomic resolvers rather than harmonizing synonyms	A general synonym-harmonization step could shift counts for many species; we flag this as a known limitation instead. The genus route (E1) already rescues most such cases — e.g. we retain *Alnus alnobetula* via genus *Alnus* in Table S2 even though its species-level EcM evidence is filed under *A. crispa* — so a residual drop-out requires the genus to be absent from Table S2 as well (Detailed methods SM1).
E11	Host-range rasters	BIEN2 MaxEnt range polygons → reproject WGS84 → clip to Canada → rasterize to **0.5° grid**; three layers (host richness, host-data richness, host-data proportion)	Standardized grid for host richness and data-coverage comparison (Detailed methods SM1).

1.1.6 Table A6. Shortfall-analysis decisions (09–21)

#	Shortfall	Decision	Choice	Rationale
F1	Linnean	Richness units	Report unique UNITE SH codes, genera, and named species (excluding `_sp`)	SH codes are the primary unit: consistent OTU applicable to both sources, and many EcM taxa are unnamed (Detailed methods SM1).
F2	Linnean	Dark fraction	Proportion of detected SH codes lacking a species-level UNITE name (combined and per source)	Operational definition of the "dark fraction" (Detailed methods SM1).
F3	Linnean	van Galen comparison	Filter van Galen per-sample table to `Country == Canada` and `biome {forest, tundra}`; report **unweighted mean** of per-sample unassigned proportion, as a non-equivalent reference only	Subset most comparable to our sampling; explicitly flagged as non-equivalent (different taxonomic units: OTU vs SH) (Detailed methods SM1).

#	Shortfall	Decision	Choice	Rationale
F4	Linnean	GBIF vouchered complement	GBIF Kingdom Fungi, Canada, basis `PRESERVED_SPECIMEN/LIVING_SPECIMEN`, geospatial issues excluded, filtered to species-rank records in EcM genera (FungalTraits)	Vouchered specimen complement to `OBSERVATION`, archived download key for reproducibility (retrieved 6 Mar 2026) (Detailed methods SM1).
F5	Wallacean	Occurrence levels	Count occurrences per taxon at SH, genus, and named-species levels	Multiple taxonomic resolutions of the shortfall (Detailed methods SM1).
F6	Wallacean	Reporting thresholds	Report proportion of taxa known from a single occurrence and from `< 30` occurrences (Canada and global); tabulate common minimum-occurrence SDM thresholds	Straightforward shortfall measures; 30 is a common SDM minimum-sample rule of thumb (Detailed methods SM1).
F7	Wallacean	Data source	Occurrence/occupancy counts drawn from **GlobalFungi only**	GlobalFungi provides the structured sampling frame required for meaningful occupancy comparison (Detailed methods SM1).

#	Shortfall	Decision	Choice	Rationale
F8	Wallacean	Global comparator	Map each Canadian species/SH to all globally-known SH codes of that identity, matched against the full global GF v5 SH matrix under the same quality filters	Consistent global denominator (Detailed methods SM1).
F9	Prestonian	Source & filters	BioTIME (downloaded 5 Mar 2026): studies with at least 1 Canadian location, at least 2 time points, fungi-consistent organism type	Minimal criteria for a fungal time series in Canada (Detailed methods SM1).
F10	Prestonian	Shortfall metric	% of named Canadian EcM species lacking any abundance time series **anywhere on Earth**	Operational definition (Detailed methods SM1).
F11	Darwinian	Source	JGI MycoCosm "All Fungi" listing (retrieved 5 Mar 2026 by copying the page; bulk export required login and was non-functional)	Only complete accessible listing at the time; fungi-only page needs no extra filtering (Detailed methods SM1).

#	Shortfall	Decision	Choice	Rationale
F12	Darwinian	Matching	**Exact** set-membership match after normalization: genus = first token of MycoCosm `taxon_name`, species = first two tokens, both lower-cased/trimmed; matched against lower-cased `emf` genera and named species (placeholder epithets sp/spp/cf/aff excluded)	We use no fuzzy/edit-distance matching or tolerance (`15_darwinian.R`). Limitation: exact matching misses cross-database synonyms and spelling variants, which slightly *over*-states the shortfall.
F13	Darwinian	Shortfall metric	% of named Canadian EcM species absent from MycoCosm	Operational definition (Detailed methods SM1).

#	Shortfall	Decision	Choice	Rationale
F14	Raunkiæran	Traits assessed	Six FungalTraits columns: 3 morphological (fruitbody type, hymenium type, exploration type) + 3 ecological (secondary lifestyle, endophytic-interaction capability, host specificity)	Columns carrying information meaningful for EcM fungi (Detailed methods SM1).
F15	Raunkiæran	Excluded trait	Growth form excluded	"Filamentous mycelium" for all EcM genera — uninformative (Detailed methods SM1).
F16	Raunkiæran	Coverage definition	Proportion of genus × trait cells non-missing and not literal "unknown"; secondary lifestyle & endophytic capability treated as a **lower bound**	Blanks in those two traits are ambiguous between "not assessed" and "trait absent" (Detailed methods SM1).

#	Shortfall	Decision	Choice	Rationale
F17	Hutchinsonian	Ecozone unit	Canada's 15 terrestrial ecozones (National Ecological Framework); report number/proportion sampled across a range of site thresholds	Standard national ecozone framework; no ecozones excluded (all contain predicted host habitat) (Detailed methods SM1).
F18	Hutchinsonian	Offshore points	Points just offshore of the ecoregion coastline snapped to nearest ecozone within **5 km**	Recover coastal samples misplaced by coastline generalization; ecozone areas use the same polygon basis (Detailed methods SM1, Table S1).
F19	Hutchinsonian	Climate variables	WorldClim 2.1 BIO1 (MAT) and BIO12 (MAP)	Two axes defining Canadian climate space (Detailed methods SM1).
F20	Hutchinsonian	Climate grid	Aggregate WorldClim 30 to ~10 (`AGG_FACTOR = 20`)	Match the spatial grain of the source workflow (Hébert et al. (2026)) (`17_hutchinsonian.R`).

#	Shortfall	Decision	Choice	Rationale
F21	Hutchinsonian	Climate binning	50 × 50 equal-width grid of temperature × precipitation; cell "sampled" if at least 1 EcM sample; coverage = sampled cells / Canada cells per zone	Coverage definition adapted from Hébert et al. (2026) (Detailed methods SM1, `17_hutchinsonian.R`).
F22	Eltonian	Sample restriction (Canadian scope)	Host associations restricted to **root-derived evidence in both databases**, by two mechanisms: GlobalFungi `sample_type == "root"`; GenBank `isolation_source` matching the root keyword rule (`root`, `mycorrhiz`, whole-word `ecm`)	Direct root extractions give strongest evidence of mycorrhizal association; bulk soil indicates only spatial proximity (Detailed methods SM1). Revised 2026-08: the submitted version applied the root restriction to GlobalFungi only and used the GenBank host field whatever the tissue, which held the two sources to different evidential standards.

#	Shortfall	Decision	Choice	Rationale
F22a	Eltonian	Records excluded by the root-evidence rule	Removed from the association analyses only; retained in the taxonomic, geographic, climatic and trait assessments	Tissue provenance bears on whether a fungus–plant pair is real evidence of symbiosis; it does not bear on whether the fungus was detected in Canada (`18_eltonian.R`, Step A2).
F22b	Eltonian	Multi-species host labels	Split on commas and semicolons; every named plant retained as a candidate host, rather than keeping only the first	`canonicalize_host()` keeps a string's first two tokens, so before this revision every plant after the first in a list was silently discarded (`18_eltonian.R`, Step A2b).

#	Shortfall	Decision	Choice	Rationale
F22c	Eltonian	Confidence flag on host associations	"Unambiguous" if the source record named exactly one candidate plant (counted across both GlobalFungi plant fields together), "ambiguous" otherwise; a pair is unambiguous if any supporting record was. All pairs reported; unambiguous-only results given as a sensitivity check	A fungus detected in a sample listing eight plants is weaker evidence than one detected where a single plant was named; flagging rather than discarding keeps the evidence and states its quality (`eltonian_pair_confidence.csv`, `eltonian_sensitivity_confidence.csv`).
F22d	Eltonian	Host-name inventory vs. association analysis	Tables S2 and S3 built from every recorded host name, whatever the tissue; all association results built from the root-derived subset	The two answer different questions — metadata quality versus available evidence. One table is built and filtered, so they cannot drift apart (`18_eltonian.R`, Step A2).
F23	Eltonian	Denominator	Native Canadian EcM host species list (Table 7) as denominator of potential interactions	Fixed, defensible universe of possible hosts (Detailed methods SM1).

#	Shortfall	Decision	Choice	Rationale
F24	Eltonian	Matrices & metrics	Interaction matrices at SH, named-species, genus levels; report % hosts with at least 1 association, matrix fill rate, and % pairs from a single record	Multiple resolutions and a single-record fragility check (Detailed methods SM1).
F25	Eltonian	Global scope retrieval	GenBank global query retrieves **all matching records per genus** (no cap), paginated via the NCBI history server	No truncation of heavily sequenced genera; a sensitivity check confirmed the earlier 2,000-record cap materially affected only a few macrofungal genera, so we removed the cap (`18_eltonian.R`).

#	Shortfall	Decision	Choice	Rationale
F25a	Eltonian	Interaction-matrix denominator	Report the full host × fungal-species grid, framed as a *metaweb* (the set of conceivable pairings), rather than a geographically filtered denominator	A geographic filter was built and tested (`21_geographic_possibility.R`). It is reported as a sensitivity check, not adopted: the result depends strongly on the spatial grain, is governed by survey effort rather than by the fungi's true ranges, removes ~44% of its cells because of missing data rather than geography, and would make the shortfall look less severe (Detailed methods SM1).
F26	Sampling maps	Sampling unit (Fig S4)	One distinct raw lat/lon per record	Location-level depiction of sampling (Detailed methods SM1, Fig S4).

#	Shortfall	Decision	Choice	Rationale
F27	Depth discard	Metric (Fig S3)	Per Canadian GF soil sample, % of total sequencing depth discarded by EcM-genus filtering, among samples with at least 1 EcM read	Quantifies the data cost of restricting to EcM genera (Detailed methods SM1, `20_depth_discard.R`).

1.2 PART B — Decisions outside our control (upstream datasets)

These decisions are embedded in the external datasets and tools we use. We report them as fully as the sources document them; where a source's internal choices are not public in detail, that limitation is itself noted.

1.2.1 Table B1. GlobalFungi v5 (Větrovský et al. (2020))

#	Upstream decision	What we know	Note
U1	Sample inclusion & curation	GlobalFungi aggregates published community-survey amplicon studies, re-processed through its own consistent pipeline (documented at globalfungi.com)	Studies use varied sequencing platforms/protocols; predominantly soil samples (Detailed methods SM1).

#	Upstream decision	What we know	Note
U2	SH pre-assignment	SH codes pre-assigned against a specific UNITE release (the build we pin to; see Supplemental Materials SM1)	We inherit GlobalFungi's SH assignment for its records rather than re-calling them.
U3	SH-assignment identity criterion	Documented BLASTn criterion at ~98.5% identity	We match this on the GenBank side (B5).
U4	Read processing / quality filtering	Internal to the GlobalFungi pipeline (denoising, chimera removal, abundance tabulation)	Details not re-implemented or independently audited by us.

1.2.2 Table B2. UNITE (Nilsson et al. (2019))

#	Upstream decision	What we know	Note
U5	SH clustering	Dynamic clustering thresholds define SH codes; identifiers renumbered between builds	Motivates our build-pinning (Supplemental Materials SM1).
U6	Representative sequence & taxonomy	One representative sequence and one taxonomy string per SH	We read taxonomy directly from these (Supplemental Materials SM1).
U7	Dark-taxon naming	Placeholder names ending `_sp` for unnamed lineages	We follow UNITE's convention in tie resolution (B6) and dark-fraction definition (F2).

1.2.3 Table B3. FungalTraits (Põlme et al. (2020))

#	Upstream decision	What we know	Note
U8	Trait scoring grain	Genus-level trait compilation	Constrains our EcM assignment and trait coverage to genus resolution (D3, F14).
U9	Primary-lifestyle categories	Includes `ectomycorrhizal` as a primary lifestyle	Basis of our EcM guild filter (D3).
U10	Blank vs "unknown" semantics	Blanks in some trait columns are ambiguous between "not assessed" and "trait absent"	Forces our lower-bound coverage treatment for two traits (F16).

1.2.4 Table B4. FungalRoot (Soudzilovskaia et al. (2020))

#	Upstream decision	What we know	Note
U11	Mycorrhiza-type labels	FungalRoot's occurrence-level label vocabulary (incl. ambiguous `EcM, AM undetermined`, `EcM,AM`)	Drives our qualifying/excluded-label choices (E2–E4).
U12	Genus-level recommendation (Table S2)	Genus calls based on at least 67% species-level consistency	Basis of our genus-level host route (E5).
U13	Underlying GBIF taxonomy	Occurrence records resolved against the GBIF backbone	Independent from BIEN resolvers → unresolved synonymy drop-outs (E9).

1.2.5 Table B5. BIEN / BIEN2 (Maitner et al. (2018); Moulatlet et al. (2025))

#	Upstream decision	What we know	Note
U14	Occurrence name-scrubbing	BIEN applies its own name resolution	Independent resolver → synonymy drop-outs (E9).
U15	Native Status Resolver	BIEN's algorithm classifies native vs introduced	We accept its native calls (E8).
U16	BIEN2 range models	MaxEnt-based modelled range polygons with fixed modelling settings (model name encodes settings)	We take ranges as given, reproject, and clip; we do not re-fit SDMs (E11).

1.2.6 Table B6. van Galen et al. (2025) dark-taxa data

#	Upstream decision	What we know	Note
U17	Dark-OTU definition	Proportion of EcM OTUs unassigned to species level, per GF v5 sample	Different taxonomic unit (OTU) from our SH-level dark fraction — comparison is non-equivalent (F3).
U18	Raster modelling	`percentage_dark_taxa` band from their modelled global layers	Used as-is for the Figure S1 map (`10_dark_diversity.R`).

1.2.7 Table B7. BioTIME (Dornelas et al. (2018))

#	Upstream decision	What we know	Note
U19	Study inclusion & structure	BioTIME's own criteria for what constitutes an included biodiversity time series	We filter within it (at least 1 Canadian location, at least 2 time points, fungi) but inherit its study boundaries (F9).

1.2.8 Table B8. JGI MycoCosm

#	Upstream decision	What we know	Note
U20	Genome inclusion	Which fungal genomes JGI has sequenced/portalized	Defines the Darwinian denominator; reflects JGI project priorities, not random sampling (F11).
U21	Listing completeness at retrieval	Page-copy snapshot on 5 Mar 2026 (bulk export non-functional)	Snapshot may lag the live portal; retrieval method documented (F11).

1.2.9 Table B9. WorldClim 2.1, GADM, Natural Earth, ecoregions, GBIF

#	Upstream decision	What we know	Note
U22	WorldClim interpolation	Climate surfaces interpolated from station data at 30 (Fick and Hijmans (2017))	We aggregate to ~10 (F20); interpolation error inherited.

#	Upstream decision	What we know	Note
U23	Administrative boundaries	GADM / Natural Earth boundary generalization	Affects coordinate validation and offshore snapping (D1, F18).
U24	Ecozone delineation	National Ecological Framework for Canada polygon definitions	Fixed unit for ecozone coverage (F17).
U25	GBIF specimen records	Basis-of-record labelling, georeferencing, and dataset coverage vary by publisher	We exclude geospatial-issue records but inherit publisher labelling (F4).

References

Dornelas M, Antão LH, Moyes F, Bates AE, Magurran AE, et al. 2018. BioTIME: A database of biodiversity time series for the Anthropocene. Global Ecology and Biogeography, 27(7): 760–786. https://doi.org/10.1111/geb.12729.

Fick SE, and Hijmans RJ. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37(12): 4302–4315. https://doi.org/10.1002/joc.5086.

Hébert K et al. 2026. Blitz the Gap: A nation-wide effort to guide citizen science toward the needs of biodiversity science.

Heeger F et al. 2018. Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. Molecular Ecology Resources, 18(6): 1500–1514. https://doi.org/10.1111/1755-0998.12937.

Maitner BS et al. 2018. The BIEN R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. Methods in Ecology and Evolution, 9(2): 373–379. https://doi.org/10.1111/2041-210X.12861.

Moulatlet GM et al. 2025. General laws of biodiversity: Climatic niches predict plant range size and ecological dominance globally. Proceedings of the National Academy of Sciences, 122(46): e2517585122. https://doi.org/10.1073/pnas.2517585122.

Nilsson RH et al. 2019. The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research, 47(D259–D264): 259–264. https://doi.org/10.1093/nar/gky1022.

Põlme S et al. 2020. FungalTraits: A user-friendly traits database of fungi and fungus-like stramenopiles. Fungal Diversity, 105(1): 1–16. https://doi.org/10.1007/s13225-020-00466-2.

Ryberg M, and Nilsson RH. 2018. New light on names and naming of dark taxa. MycoKeys, 30: 31–39. https://doi.org/10.3897/mycokeys.30.24376.

Soudzilovskaia NA et al. 2020. FungalRoot: Global online database of plant mycorrhizal associations. New Phytologist, 227(3): 955–966. https://doi.org/10.1111/nph.16569.

van Galen LG et al. 2025. The biogeography and conservation of Earth's "dark" ectomycorrhizal fungi. Current Biology, 35(11): R563–R574. https://doi.org/10.1016/j.cub.2025.03.079.

Větrovský T et al. 2020. GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. Scientific Data, 7(1): 228. https://doi.org/10.1038/s41597-020-0567-7.