ARTICLES
https://doi.org/10.1038/s41559-019-0975-2
nature ecology & evolution

Host availability drives distributions of fungal endophytes in the imperilled boreal realm

Jana M. U'Ren[1], François Lutzoni[2], Jolanta Miadlikowska[2], Naupaka B. Zimmerman[3], Ignazio Carbone[4], Georgiana May[5] and A. Elizabeth Arnold[6,7]*

Boreal forests represent the world's largest terrestrial biome and provide ecosystem services of global importance. Highly imperilled by climate change, these forests host Earth's greatest phylogenetic diversity of endophytes, a hyperdiverse group of symbionts that are defined by their occurrence within living, symptomless plant and lichen tissues. Endophytes shape the ecological and evolutionary trajectories of plants and are therefore key to the function and resilience of terrestrial ecosystems. A critical step in linking the ecological functions of endophytes with those of their hosts is to understand the distributions of these symbionts at the global scale; however, turnover in host taxa with geography and climate can confound insights into endophyte biogeography. As a result, global drivers of endophyte diversity and distributions are not known. Here, we leverage sampling from phylogenetically diverse boreal plants and lichens across North America and Eurasia to show that host filtering in distinctive environments, rather than turnover with geographical or environmental distance, is the main determinant of the community composition and diversity of endophytes. We reveal the distinctiveness of boreal endophytes relative to soil fungi worldwide and endophytes from diverse temperate biomes, highlighting a high degree of global endemism. Overall, the distributions of endophytes are directly linked to the availability of compatible hosts, highlighting the role of biotic interactions in shaping fungal communities across large spatial scales, and the threat that climate change poses to biological diversity and function in the imperilled boreal realm.

As the world's largest terrestrial biome, boreal forests span more than 11% of Earth's land area and comprise around 30% of the global forest cover[1]. Boreal forests exert the greatest biogeophysical effects on mean global temperature and harbour a disproportionately high level of carbon in soil, which—when combined with boreal vegetation—makes up around 50% of the planet's atmospheric carbon[2]. By 2100, warming due to climate change is expected to have a profound effect on biodiversity and species composition in boreal forests[3], yielding substantial downstream effects on the net carbon balance and climate feedback effects driven by these high-latitude ecosystems[4–6].

Plant-associated microbial communities are increasingly recognized for their potential to facilitate rapid acclimatization of plants to novel stressors, especially within threatened biomes[7,8]. Soilborne and root-associated fungi are critical to nutrient cycling, soil dynamics, and ecosystem productivity and resilience in boreal ecosystems[9,10]. Long under-studied owing to their cryptic occurrence in healthy aboveground tissues, fungal endophytes that occur within photosynthetic tissues of plants and in association with photosynthetic partners in lichens[11] also are key players in host health, productivity and stress mitigation[12–15]. Endophytes originated contemporaneously with the origin of land plants[16], and comparative studies reveal that they reach their greatest phylogenetic diversity in boreal forests, exceeding their phylogenetic diversity in tropical regions[17]. In highly imperilled boreal forests, understanding the distributions of endophytes is a critical first step in linking their ecological functions with those of their hosts, and is key to interpreting the resilience of ecosystems such as forests to environmental change[2].

The majority of fungal endophytes are transmitted horizontally, and over broad spatial scales their distributions generally reflect abiotic factors, such as climate or geographical distance[18,19], similar to free-living fungi in soil[20]. However, host communities shift in composition with geography and climate, often confounding inferences about symbiont biogeography[18]. As a result, there is a need to disentangle deterministic processes, such as host and environmental filtering, from neutral processes, such as dispersal and drift, as drivers of endophyte diversity and distributions at a circumglobal scale. Boreal forests represent a unique opportunity to disentangle these factors because of their broad consistency in vegetation types, climate and phylogenetic composition of plant and lichen communities across continental to intercontinental scales[1].

We examined endophyte communities by culture-based sampling and culture-independent next-generation sequencing (NGS) of 498 individual plant-host and lichen-host collections that were newly obtained in seven sites in North America and Eurasia that, together, circumscribe the global boreal belt (Fig. 1a and Supplementary Table 1). At each site, we collected photosynthetic tissues from living asymptomatic plants, representing Magnoliophyta, Pinophyta, Monilophyta, Lycopodiophyta and Bryophyta, as well as lichens that comprised fungal mycobionts with Cyanobacteria, Chlorophyta or both photobionts on soil and moss, rock, bark or dead wood (Fig. 1b, Supplementary Fig. 1, and Supplementary Tables 2 and 3). Overall, our sampling included at least 60 individual plant and lichen collections per site (60–105 collections) and an average of 19 host genera per site (17–23 genera; Fig. 1). Geographical distances between individual host

[1]Department of Biosystems Engineering and BIO5 Institute, University of Arizona, Tucson, AZ, USA. [2]Department of Biology, Duke University, Durham, NC, USA. [3]Department of Biology, University of San Francisco, San Francisco, CA, USA. [4]Center for Integrated Fungal Research, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA. [5]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN, USA. [6]School of Plant Sciences, University of Arizona, Tucson, AZ, USA. [7]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA. *e-mail: arnold@ag.arizona.edu

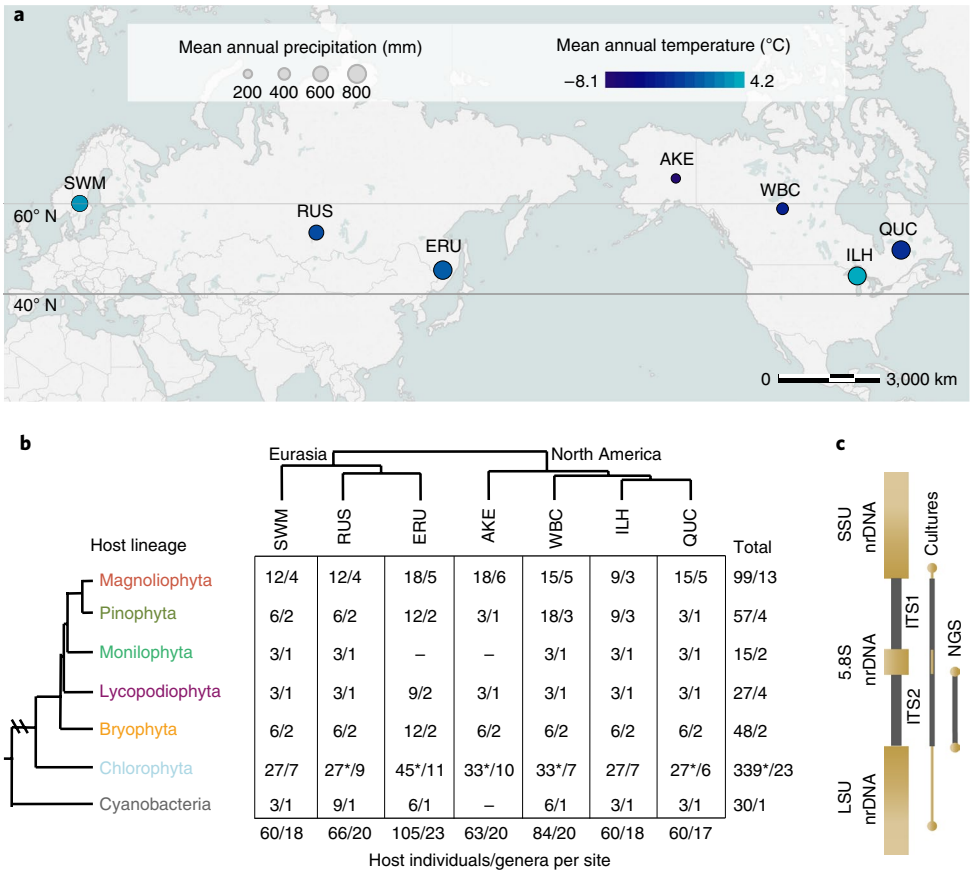

Fig. 1 | Geographical location, climate and host information for 498 individual host collections sampled for endophytes at seven boreal sites. **a**, Sampling sites: Mora, Sweden (SWM); Stolby Reserve, Krasnoyarsk Territory, Russia (RUS); Bol'shekhekhtsirskii Reserve, Khabarovsk Territory, Russia (ERU); Eagle Summit, Alaska, USA (AKE); Wood Buffalo National Park, Canada (WBC); Ives Lake Field Station, Huron Mountains, Michigan, USA (ILH); Québec, Canada (QUC). Map source: Base map © Mapbox © OpenStreetMap. **b**, The number of host individuals and host genera collected at each site, depicted with relative geographical distances among sites (top dendrogram) and relationships of photobiont host lineages (left). Lichen photobionts include Chlorophyta or Cyanobacteria, which can occur alone within thalli (that is, in chlorolichens or cyanolichens, respectively) or together in one thallus (that is, tripartite lichens (asterisks)). Tripartite lichens are counted above only once as Chlorophyta because the lichen thalli that we collected were dominated (in area or volume) by the algal photobionts. **c**, Fungal barcode locus: nuclear ribosomal ITS and 5.8S gene, sequenced for cultures with a portion of the nuclear ribosomal LSU (ITS nrDNA-partial LSU nrDNA). NGS data represent the ITS2 nrDNA region.

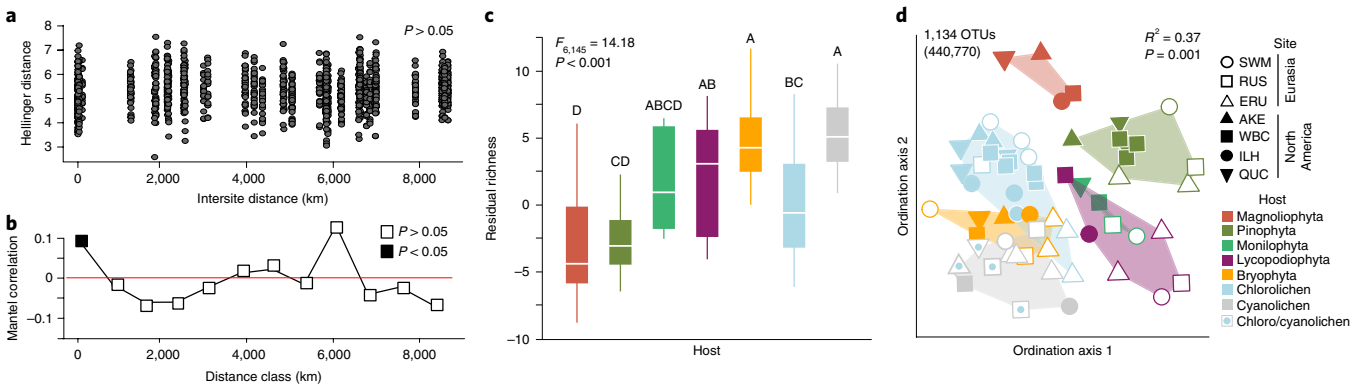

Fig. 2 | Host identity structures endophyte communities at a circumboreal scale. **a**, We observed no evidence of distance decay in endophyte community similarity. **b**, A Mantel correlogram as a function of geographical distance classes among seven sites shows a lack of geographical autocorrelation. **c**, A quantile box plot illustrates variation in richness among host lineages. Statistical analyses were performed using analysis of variance with post hoc Tukey's HSD (different letters indicate significant differences in endophyte richness among host lineages). **d**, NMDS for one host genus per lineage (colours) at a circumboreal scale reveals differences in endophyte communities among host genera, including those host genera sampled in at least four sites (shapes): *Rhododendron* (Magnoliophyta), *Picea* (Pinophyta), *Equisetum* (Monilophyta), *Lycopodium* (Lycopodiophyta), *Pleurozium* (Bryophyta), *Cladonia* (chlorolichen) and *Peltigera* (cyanolichens and tripartite lichens). Statistics reflect host × site interaction in permutational multivariate analysis of variance after combining data for three individual replicates for each host species per site (see Methods; Supplementary Table 6).

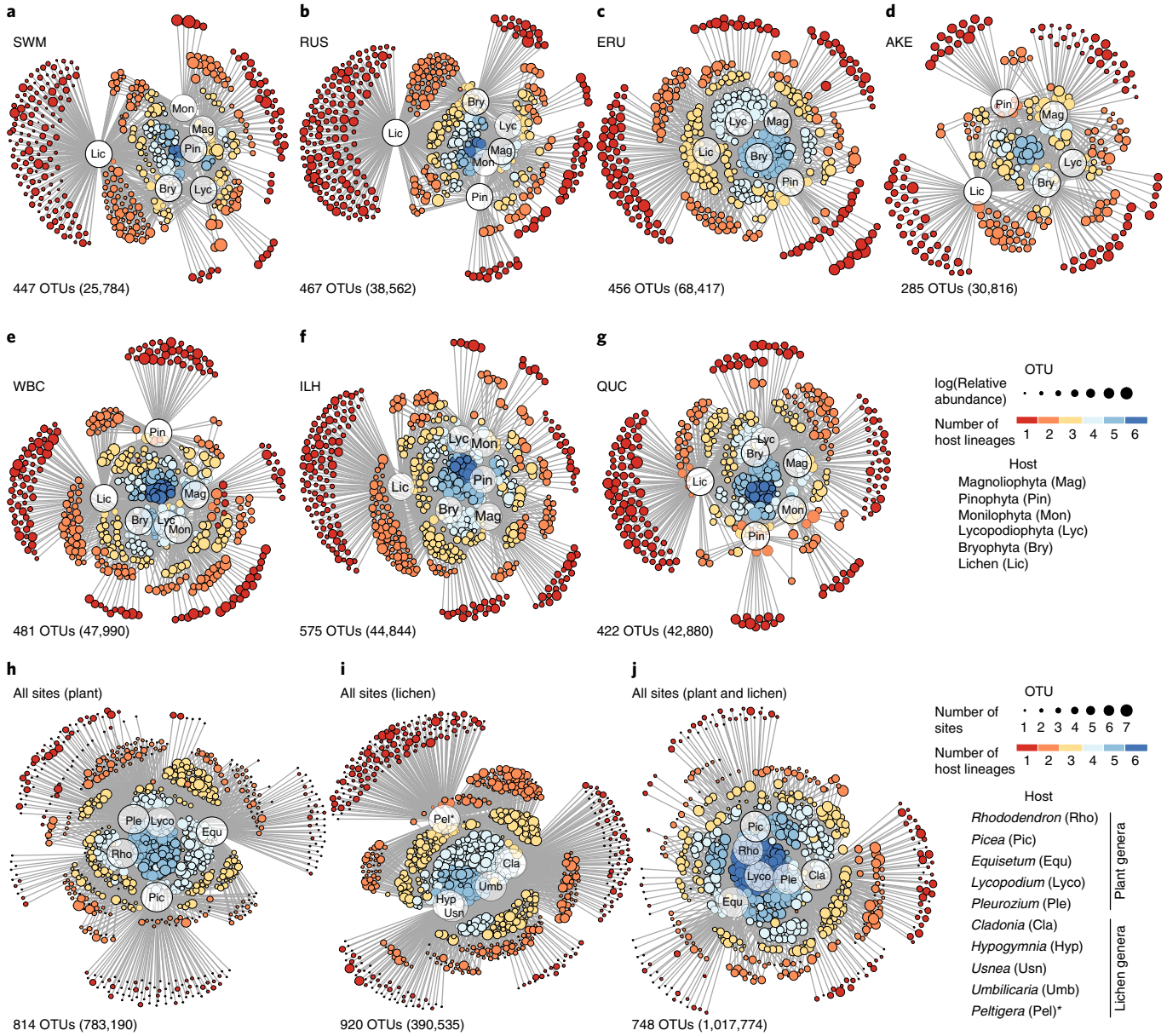

Fig. 3 | Networks reveal host affiliations of endophyte OTUs at local and circumboreal scales. a–g, Networks by site, with node diameter proportional to $\log_{10}$-transformed read abundance. Colour indicates the number of host lineages in which an OTU was observed. OTU richness and read depth are shown for each host lineage per site. **h–j**, Networks for circumboreal datasets; node diameter is proportional to the number of sites in which the OTU was observed. Host lineages are represented by a single host genus sampled in at least four sites. Asterisks indicate cyanolichens. For **a–j**, nodes represent OTUs. Edges connect OTUs to host lineage(s) in which they were found.

collections ranged from local (<1–100 m) to global scales (up to 8,676 km; Fig. 1a). Host tissues were surface-sterilized and cut into 2 mm^2 fragments[18] for culturing and NGS. In total, we prepared more than 46,000 fragments to isolate endophytes and an equivalent quantity for NGS (Supplementary Tables 2 and 3). We obtained 11,975 endophyte isolates in culture and used the Sanger platform to sequence the fungal internal transcribed spacer nuclear ribosomal DNA (ITS nrDNA) barcode locus[21] and a portion of the adjacent, phylogenetically informative large subunit (LSU) nrDNA region for each isolate (Fig. 1c and Supplementary Tables 2 and 3). NGS analysis of host tissues generated more than 6 million quality-filtered ITS2 nrDNA sequences of endophytes (Fig. 1c and Supplementary Tables 2 and 3).

Results and discussion

Together, the culture-based and NGS datasets included more than 6,000 operational taxonomic units (OTUs) in five fungal phyla, including a minimum of seven classes of Ascomycota and seven classes of Basidiomycota (Supplementary Fig. 2). Endophyte richness values inferred by culturing and NGS were correlated positively, independent of host lineage or sequencing depth (Supplementary Fig. 3). Overall, richness based on NGS was approximately fifteenfold greater than that inferred by culturing from the equivalent quantity of host tissue (Supplementary Tables 2 and 3). When NGS data were subsampled to match the number of sequenced cultures, NGS provided an approximately fivefold increase in richness relative to culturing (NGS, 1,466.5 ± 21.3 OTUs; culturing,

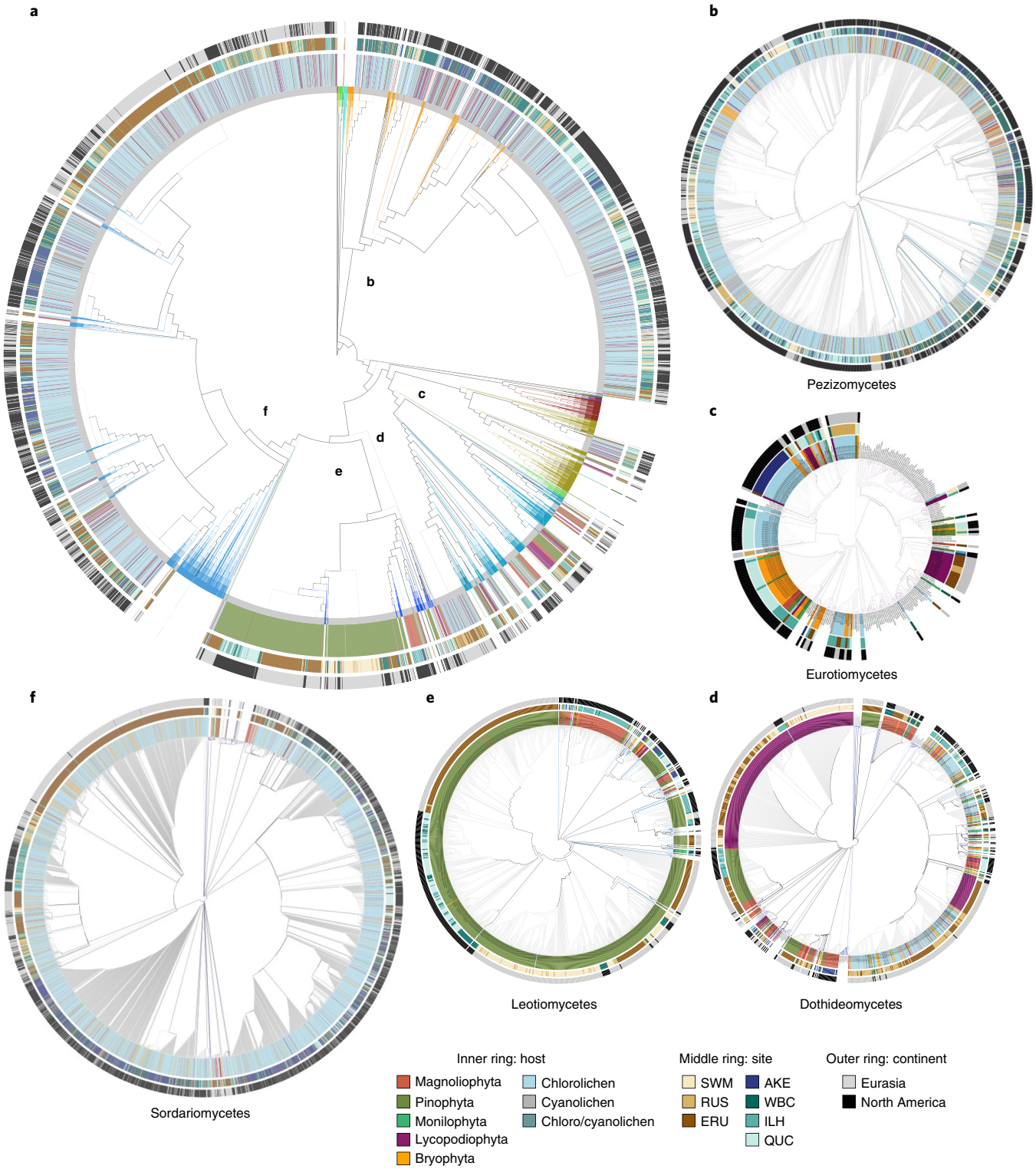

Fig. 4 | Evolutionary context of endophyte–host associations revealed by phylogenetic analyses of the most species-rich fungal phylum (Ascomycota). Phylogenetic placement of endophytes was inferred in T-BAS[30]. Trees show endophyte isolates, which were obtained by culturing, with rings of metadata (host, site, continent) indicated by colours. Reference taxa are shown with coloured branches and no metadata[30]. **a**, Pezizomycotina, the largest subphylum of Ascomycota, with letters corresponding to panels **b**–**f**, which represent the most endophyte-rich classes of Pezizomycotina. The diameter of each circular tree represents the relative abundance of each focal class. Lichens have photobionts as described in Fig. 1. Support values are shown in Supplementary Figs. 12–16.

315.0 ± 7.0 OTUs). Culturing and NGS recovered the same classes and orders of Ascomycota, albeit in different proportions, whereas NGS recovered a higher diversity of Basidiomycota (Supplementary Fig. 2 and Supplementary Methods). Sampling was sufficient for ecological inference (Supplementary Fig. 4) and repeated sampling at a focal site after 3 yr showed that a single sampling event was

representative of the local endophyte community over the timescale of our study (Supplementary Fig. 5).

Comparison of the entire dataset with more than 44,000 OTUs observed in global surveys of soil fungi[22], including fungi from boreal soils, revealed that boreal endophytes were strikingly distinct. Only 1.5% of OTUs found in the global soil dataset were observed here as endophytes[22] (Supplementary Table 4). Similarly, only 2.5% of OTUs observed here were found in comparable surveys of endophytes from the temperate zone (12% when data were restricted to only cultures, as in previous studies[18]; Supplementary Table 4).

These findings underscore the tremendous richness of boreal endophytes and the distinct niche they occupy as symbionts. As such, we evaluated the importance of host identity, climate and geographical distance in structuring endophyte assemblages at local to circumboreal scales. Within each site, host identity was the major predictor of endophyte community structure (Supplementary Fig. 1). Host genus explained an average of 58% of the variation in endophyte community composition within sites (51–68%; Supplementary Fig. 1). As for soil fungi[22,23], endophyte richness was correlated positively with mean annual precipitation (MAP; Supplementary Table 5). For boreal endophytes, however, host lineage had greater explanatory power than MAP in our models (Supplementary Table 5 and Supplementary Fig. 6).

At the circumboreal scale, we predicted that dissimilarity of endophyte assemblages would correlate positively with geographical distance, consistent with distance decay[24]. However, dissimilarity of endophyte assemblages could not be explained by geographical distance (Fig. 2a,b). Instead, host effects persisted at the global scale (Fig. 2c,d), reflecting the positive correlation between community dissimilarity of endophytes and the genetic distance between host taxa across the circumboreal belt (Mantel test: $r=0.20$, $P<0.0001$; see also ref. [25] for similar correlations for root-associated bacteria, but at local scales). These host effects were modulated by site-specific factors, the importance of which varied among host lineages (Supplementary Fig. 7 and Supplementary Table 6). Thus, assemblages of boreal endophytes appear to largely reflect biotic filtering by hosts in the context of distinctive environments, microclimates or historical artefacts of host distributions[26], rather than turnover due to inter-site distance alone. Accordingly, the slopes of species–area relationships for boreal endophytes are steep regardless of geographical scale (Supplementary Fig. 8 and Supplementary Table 7).

For horizontally transmitted symbionts, host colonization requires both dispersal to the host and symbiotic establishment. Endophyte OTUs with wide host ranges might be predicted to have large geographical ranges owing to the widespread availability of suitable partners[27]. To test this prediction, we used networks to visualize the associations of endophyte OTUs with hosts at local and circumboreal scales (Fig. 3). Even when analyses were restricted to the most common OTUs, an average of 64% of OTUs were affiliated with members of only one host lineage in each site (Fig. 3a–g). The number of host lineages in which an OTU was found was a poor predictor of OTU abundance, suggesting that OTUs were not designated inappropriately as specialists simply because they were rare (Supplementary Fig. 9 and Supplementary Methods). When scaled to the circumboreal level, an average of 24% of the most common OTUs still associated with members of only one host lineage (Fig. 3h–j). The remaining OTUs seem to be host generalists with wider geographical distributions than the more locally restricted specialists (Supplementary Fig. 10). Although it is possible that the apparent generalists contain cryptic species with narrower distributions[28], haplotype analysis of sequences representing the most widespread generalist OTU reveals a global distribution of the most abundant amplicon sequence variants (ASVs; Supplementary Fig. 11). Geographically restricted and specialist OTUs represent diverse genera with different spore sizes and discharge methods (including endophytes that are closely related to plant pathogens with transoceanic dispersal[29]), such that dispersal limitation alone cannot explain their limited distributions. The availability of suitable hosts likely limits the geographical distributions of specialists and drives the high global richness of endophytes at a circumboreal scale.

Acknowledging the evolutionary relatedness among OTUs provides an important framework for understanding ecological patterns. At present, the relatively short sequences generated by NGS for fungi usually cannot be placed reliably in community-scale phylogenetic analyses[21,30]. The endophyte OTUs that we isolated in culture were a representative subset of abundant OTUs obtained by NGS from the same host material (Supplementary Methods) but, in contrast to the short sequences obtained by NGS, were represented by longer sequencing reads containing regions that are informative for phylogenetic placement[30]. By placing these cultured endophytes in a robust phylogenetic framework for the first time, we detected distinctive evolutionary trajectories in each focal class of the most prevalent phylum in both of the culture-based and culture-independent datasets (that is, Ascomycota; Fig. 4 and Supplementary Figs. 12–16). We observed a relatively wide host generalism and broad geographical distributions of endophyte-dominated clades in Sordariomycetes and Pezizomycetes, which affiliate especially frequently with lichens (Supplementary Fig. 2). In contrast, endophytes in classes such as Dothideomycetes and Leotiomycetes often had narrower host and geographical distributions, and were observed more frequently in plants (Supplementary Fig. 2).

Fungal endophytes influence the functional traits, ecological dynamics and evolutionary trajectories of their hosts, and are therefore fundamentally important to the dynamics and resilience of plant communities under climate stress[14]. Experimental studies reveal direct sensitivity of boreal endophytes to warming and suggest altered functional roles with climate change[31,32]. Our results suggest that endophytes of boreal plants and lichens are distinctive, hyperdiverse and distributed in a manner that reflects the presence of compatible hosts at local to circumglobal scales. Thus, shifts in climate that lead to local and regional extirpation of plants and lichens[4] are likely to result in the rapid loss of endophyte diversity locally. As a consequence, boreal plant and lichen communities globally may face a loss of symbiont-conferred resilience—a change that is detrimental to their continued persistence in the increasingly imperilled boreal realm.

Methods

Field collections. We collected fresh, photosynthetic tissues of a diverse range of plants and lichens in seven sites across North America and Eurasia (Fig. 1 and Supplementary Table 1). Climate data were obtained from the WorldClim database (www.worldclim.org) at a resolution of 30 arcsec. There was no evidence of recent fire in any focal site (on the basis of tree cores, interviews with forestry agents, forestry data and observations of fire damage such as charcoal, scarring and related indicators). Field collections were conducted at the height of the growing season from 2011 to 2013 (Supplementary Table 1). At each site, we collected fresh mature asymptomatic tissues of at least ten species of plants and thalli of at least ten species of lichens (defined by mycobiont) in each of three replicate microsites as described previously[18] (Fig. 1 and Supplementary Tables 2 and 3). For each host we collected a random subset of photosynthetic tissues that, for long-lived individuals or tissues, encompassed multiple years of growth. Portable laminar-flow hoods facilitated sterile processing at remote locations, and sterile methods were used for all of the tissue processing steps described below.

Endophyte isolation, DNA extraction, PCR amplification and Sanger sequencing. Fresh tissues from each host collection were cut into 2 mm^2 segments, which were surface-sterilized as described previously[18]. We chose 96 segments randomly for endophyte isolation, and an equal number were chosen randomly for culture-independent analysis (described below)[33]. Endophytes were isolated on 2% malt-extract agar under sterile conditions[18]. Fungi that emerged from tissue pieces were vouchered in sterile water and deposited at the Robert L. Gilbertson Mycological Herbarium at the University of Arizona (Supplementary Table 8). We extracted total genomic DNA directly from each fungal isolate[34]. The nuclear ribosomal ITS and 5.8S gene (ITS nrDNA) and an adjacent portion of the nuclear ribosomal LSU (LSU nrDNA; approximately 500 bp) were PCR-amplified as a single fragment (Fig. 1c), sequenced bidirectionally and processed manually[18].

High-quality Sanger sequencing data were obtained for 10,805 out of 11,975 isolates (Supplementary Tables 2, 3 and 8).

DNA extraction, PCR amplification and Illumina sequencing. Concurrently with culturing (described above), we placed 96 surface-sterilized segments per host collection in cetrimonium bromide (CTAB) buffer (1 M Tris-HCl pH 8, 5 M NaCl, 0.5 M EDTA and 20 g CTAB)[33,35] under sterile conditions. Tubes were stored at −80 °C until the DNA was extracted. We extracted total genomic DNA using the MoBio PowerPlant Pro DNA Isolation Kit (Qiagen)[36] and amplified and sequenced the fungal ITS nrDNA locus for each sample using a dual-barcoded two-step library preparation process[37] with the primer pair ITS1F and ITS4[38,39]. We performed PCR for each sample in three replicates. Amplification was verified on 2% agarose gels that were stained with SYBR Green I (Molecular Probes, Invitrogen). The final PCR products were quantified fluorometrically, normalized and pooled in equimolar amounts. The final amplicon pool was purified with Agencourt AMPure XP beads according to the manufacturer's instructions (Beckman Coulter). A BioAnalyzer 2100 (Agilent Technologies) was used to determine the DNA concentration and fragment-size distribution of the final library before paired-end sequencing was performed using an Illumina MiSeq with the Reagent Kit v3 (2 × 300 bp).

Bioinformatics and quality control. Raw Illumina data were demultiplexed and sequences representing PhiX and a diversity shotgun library (that is, genomic DNA representing a non-fungal organism that was spiked into the run to improve cluster density during sequencing; IBEST Genomics Core, personal communication), as well as sequences containing more than one mismatch to the barcode and more than four mismatches to primers, were removed. The remaining 9,942,458 reads that corresponded to the ITS2 nrDNA region were trimmed for quality control using a cut-off length of 170 bp and a maximum error rate of 1.0 in USEARCH v.8.1.1861[40,41], resulting in 4,553,953 high-quality sequences. To combine Sanger sequences from cultures with Illumina sequences for direct comparisons, we first used ITSx v.1.0.7[42] to identify Sanger sequences that did not contain at least 50 bp of either ITS1 or ITS2 nrDNA. These sequences ($n = 86$) were removed. For the remaining 10,719 Sanger sequences (Supplementary Tables 2 and 3), all of the bases downstream of the conserved region at the start of LSU nrDNA (that is, 3′ end) were removed and the 5′ end of the sequences were trimmed to a length of 170 bp to match the exact length and start position of Illumina sequences. Sanger and Illumina sequences were dereplicated in parallel and OTUs represented by only one or two Illumina sequences (that is, singleton or doubleton OTUs) were removed[33,40].

OTU clustering and taxonomic assignments. After these filtering steps, dereplicated sequences from both the culture-based and NGS analyses were clustered into OTUs at 95% sequence similarity using the UPARSE-OTU algorithm[43] as implemented in USEARCH[40], a decision that was based on the clustering results of the mock community (Supplementary Methods). In addition to de novo chimera checking that was performed during clustering[44], representative sequences for each OTU were subjected to reference-based chimera checking using the UNITE[45] database with UCHIME[44]. Raw Illumina reads and all of the Sanger reads were mapped back to chimera-checked OTUs to construct an OTU table containing more than 6 million reads and more than 6,200 OTUs.

A representative sequence from each OTU (which was chosen to represent the most abundant sequence in the cluster) was queried first with ITSx[42] to identify and subsequently remove OTUs that lacked the ITS2 region. Sequences from the remaining OTUs were queried against the NCBI non-redundant nucleotide database (but excluding all environmental sequences) with BLASTn[46]. BLAST output was analysed using MEGAN[47] v.5.11.3 with default parameters for the lowest common ancestor. OTUs representing lichen-forming fungi (that is, the primary mycobiont; Supplementary Table 3) or plant hosts, sequences with no hits, and/or sequences not classified to fungi, were removed from the subsequent analyses. The remaining OTUs were queried against the UNITE fungal database[45] with the RDP Classifier[48] for taxonomic classification with a cut-off threshold of 80% confidence as implemented in QIIME v.1.8[49] (Supplementary Fig. 2). Analyses of the phylogenetically diverse mock community confirmed our bioinformatic methods for (1) removal of spurious OTUs resulting from erroneous sequences; (2) low prevalence of tag-switching[50] among samples (<1% of OTUs); and (3) correct estimates of species boundaries for phylogenetically diverse taxa present in the mock community (Supplementary Table 9). Representative analyses described below were repeated with data that were denoised and clustered into zero-radius OTUs (that is, zOTUs; analogous to amplicon sequence variants[51]) with the UNOISE2 algorithm[52] in USEARCH (Supplementary Methods). In this context, rare taxa were more abundant, but our main results did not differ appreciably.

Sanger sequences containing the entire ITS nrDNA-partial LSU nrDNA region were also clustered into OTUs independent of NGS reads following methods described previously[18]. We observed a significant correlation in species richness per host between OTU richness based on full-length ITS nrDNA-partial LSU nrDNA Sanger sequences and OTU richness based on trimmed ITS2 nrDNA reads (to match NGS read length, see above; Pearson correlation: $r = 0.98$, $P < 0.0001$; Supplementary Tables 2 and 3). OTUs that were designated using full-length ITS nrDNA-partial LSU nrDNA Sanger sequences were used for analyses of richness based on culturing (see below; Supplementary Tables 2 and 3).

Comparison of boreal endophytes to a global survey of fungi from soil. A representative sequence for each OTU was clustered with representative sequences for 44,563 fungal OTUs from a global survey of soil[22] at 99% ITS nrDNA sequence identity (to account for differences due to different sequencing and bioinformatic methods between studies) with UCLUST[40]. Percentage overlap was calculated as the number of boreal endophyte OTUs that clustered with a fungal OTU from soil divided by the total number of OTUs from soil fungi (Supplementary Table 4). Similar results were obtained when clustering was repeated using 97% sequence similarity (that is, 3% of soil OTUs[22] were observed here as boreal endophytes).

Comparisons of boreal endophytes to endophytes of temperate plants and lichens. We compared the overlap of boreal endophytes with endophytes from plants and lichens in a temperate semideciduous forest, temperate coniferous forest and subtropical scrub forest of North America[18], which were isolated, sequenced and analysed using the methods described here ($n = 1{,}042$ cultures; 352 OTUs). Intersite distances between boreal sites and these sites ranged from 4,452 km to more than 10,000 km. Percentage overlap was calculated as the number of OTUs that contained both boreal and temperate endophytes divided by the total OTUs for cultured endophytes only (12.0%; 63 out of 524 OTUs), as well as all of the cultures plus NGS reads (2.5%; 153 out of 6,152 OTUs; Supplementary Table 4).

Phylogenetic analyses. Phylogenetic placement of endophytes in the Ascomycota was inferred using the Tree-Based Alignment Selector Toolkit (T-BAS) v.2.1 (https://tbas.hpc.ncsu.edu/)[30] with the evolutionary placement algorithm in RAxML[53] for 10,805 cultures of boreal endophytes for which ITS nrDNA-partial LSU nrDNA sequences were obtained (Fig. 4 and Supplementary Figs. 12–16). The reference Pezizomycotina tree in T-BAS is based on six loci[30]. The settings that we used to place endophyte cultures within the reference Ascomycota tree with 5.8S nrDNA and partial LSU nrDNA sequences were as follows: UNITE filter engaged, 1.0 sequence identity, genetic distance score = 10 standard deviations, likelihood weights (fast), with the outgroup selected. Each major class of Pezizomycotina was then selected (Fig. 4) for RAxML analysis with 1,000 bootstrap replicates following realignment in MAFFT[54], and data were retained for all of the cultures (Fig. 4 and Supplementary Figs. 12–16). Haplotype network analyses for sequences of *Daldinia loculata* (see ref. [55] for phylogenetic placement) were performed in T-BAS with TCS v.1.21[56] (Supplementary Fig. 11).

Statistical analyses. We used analysis of variance to compare richness among sites and host lineages for cultures and NGS after accounting for differences in sequencing depth. Richness was defined for the analysis by calculating the residuals of OTU richness in relation to the square root of the number of reads as described previously[22] (Fig. 2 and Supplementary Fig. 6). We examined the relationship between endophyte species richness and environmental variables (mean annual temperature (MAT) and MAP), host lineage and site using linear mixed models (Supplementary Table 5). We compared total richness among host lineages and sites for both cultures and NGS data using rarefaction (Supplementary Fig. 4). We calculated OTU-richness estimates and performed rarefaction analyses using the vegan[57] package in R[58].

We used non-metric multidimensional scaling (NMDS) ordinations based on Hellinger dissimilarity to visualize fungal community structure within each site (Supplementary Fig. 1). We used all of the host collections from our main sites (Fig. 1 and Supplementary Tables 2 and 3). We used the same approach for the analysis across seven sites that span the circumboreal belt; however, for each major host lineage we used data from a single representative genus sampled in at least four sites: *Rhododendron* (Magnoliophyta), *Picea* (Pinophyta), *Equisetum* (Monilophyta), *Lycopodium* (Lycopodiophyta), *Pleurozium* (Bryophyta), *Cladonia* (chlorolichen) and *Peltigera* (cyanolichens and tripartite lichens; Fig. 2d). Read counts among samples differed by more than 2–3×; thus, to remove the effect of differential sequencing depth we rarefied the number of NGS reads per host to the lowest number of sequences following previous recommendations[59]. Owing to the preponderance of zeros in the OTU matrix, non-convergence of the ordination search and high stress values, NMDS analyses at the circumglobal scale were restricted to OTUs with more than 100 reads. Data for three collections for each host species per site were combined to enable the NMDS analysis to converge (Fig. 2d).

We used permutational multivariate analysis of variance (PERMANOVA) with the Hellinger distance metric to assess the significance of community similarity as a function of host genus and lineage (plant/algal phylum or mycobiont order) at the local scale (Supplementary Fig. 1), or as a function of host identity (that is, lineage and genus), site and/or environmental variables (MAT, MAP) at the circumglobal scale (Supplementary Table 6). In these analyses, data from multiple microsites were not combined (whereas in Fig. 2d, data from multiple microsites were combined to achieve convergence of the NMDS analysis and lower stress values; see above). Site explained a greater proportion of variation in endophyte community composition than MAP and/or MAT; therefore, we used site as an explanatory variable because it encapsulates both climate as well as other site-specific factors. PERMANOVA were implemented using the 'adonis' function in

the R library vegan[57] as described previously[60–62]. To account for the significant effect of host on endophyte community structure, analyses at the circumglobal scale were conducted with the entire dataset as well as various subsets of hosts including (1) only plants; (2) only lichens; or (3) various combinations of 16 plant and lichen genera, each of which was found in a minimum of four sites (Supplementary Table 6).

Relationship between richness estimates from cultures and NGS. We tested for a correlation between species richness as inferred from culture-free NGS and Sanger sequencing using Pearson's correlation coefficient (Supplementary Fig. 3). To account for differences in sample sizes, reads for each host species were rarefied to the lowest read depth (Supplementary Tables 2 and 3 and Supplementary Fig. 3). We examined the strength of the correlation after calculating NGS richness in two ways: (1) using similar NGS sampling depth per host species/site (see above, rarefaction); and (2) using the same number of sequences as those obtained from cultures, and focusing only on Ascomycota.

Assessment of interannual variation in endophyte communities. We compared the isolation frequency, richness and community composition of endophytes isolated in culture in one site (AKE; Supplementary Table 1) in summer of 2008 (ref. [18]) and 2011 (Supplementary Fig. 5). Isolation frequency, defined as the percentage of tissue segments containing cultivable fungi, was used as a proxy for host tissue colonization[18]. We used *t*-tests to compare isolation frequency between sampling years for plants and lichens separately (Supplementary Fig. 5). As sampling intensity was 2× greater in the second sampling year, we rarefied reads 1,000× to compare richness at the same sequencing depth (Supplementary Fig. 5). We used PERMANOVA to test for differences in endophyte community composition as a function of sampling year and visualized endophyte communities with NMDS as described above (Supplementary Fig. 5).

Spatial autocorrelation and distance decay. We computed Mantel correlograms of Hellinger community distance and intersite geographical distances (Fig. 2b) to quantify spatial autocorrelation[63]. Intersite distances were measured using the Haversine method in the R package fields[64]. Correlation coefficients were computed after 999 permutations. Relationships between community distance and intersite distances for Sanger sequences and NGS data were plotted to visualize distance decay (Fig. 2a) and Mantel tests were computed to test for a correlation. To test the significance of site and host lineage on communities while constraining variation attributable to distance alone, we used distance-based redundancy analysis (RDA) constrained by principal components of neighbour matrices (PCNM), implemented in vegan using the 'capscale' function[61,65,66]. The 'ordiR2step' function in vegan was used for forward model choice solely on adjusted R^2 and P values. RDA also was used to assess variation attributable to spatial eigenvectors alone, after accounting for host lineage and site effects.

Hierarchical clustering of endophyte communities in focal host genera. We used UPGMA average linkage clustering with Bray–Curtis dissimilarity in vegan[57] to assess the importance of site-specific factors on endophyte community composition in focal host genera. If geographical distance affects endophyte community composition, endophyte communities within a single host genus should cluster according to intersite distances (Fig. 1b). Instead, UPGMA dendrograms appear to illustrate site-specific factors, the importance of which varied among host genera (Supplementary Fig. 7).

Relationship of host genetic distance and endophyte community dissimilarity. We used Mantel tests to examine the correlation between host genetic distance and endophyte community dissimilarity. Endophyte community dissimilarity was defined with Hellinger distance (see above) and host genetic distance was estimated by analysis of sequencing data that represent the ribulose bisphosphate carboxylase large chain (*rbcL*) for plants. A lack of data for many mycobionts (that is, approximately 66% of mycobionts for locus *RPB1*) precluded a similar analysis for lichens. The matrix of host *rbcL* uncorrected pairwise distances was computed in mothur[67] using the default parameters. A Mantel test was implemented with vegan[57] in R using the Pearson correlation method and 999 permutations.

Species–area relationships. Species–area relationships were computed for Sanger sequences from cultures and NGS data on the basis of sampling area and area of photosynthetic tissues (Supplementary Fig. 8 and Supplementary Table 7). Species richness was calculated as the mean richness of all of the possible permutations at each sampling area (Supplementary Methods). For each analysis, species richness and area were $\log_{10}$-transformed before regression.

Endophyte host associations. We quantified and visualized the distribution of OTUs among major host lineages with networks constructed with the R package igraph v.0.7.1[68]. Networks were constructed for OTUs in each site (using endophytes from all of the host taxa; Supplementary Tables 2 and 3). Networks constructed at a circumglobal scale were restricted to (1) communities from a subset of ten plant genera and five lichen genera, each of which was sampled in at least four sites (Supplementary Fig. 9) or (2) endophyte communities from a representative genus for each major host lineage (Fig. 3). We used χ^2 tests to evaluate the null hypothesis that the number of host lineages used by an endophyte OTU was consistent regardless of the number of sites in which that OTU was found (that is, one site, two sites and so on). Likelihood ratio tests were used to compute the probability of obtaining, by chance alone, a χ^2 value greater than the observed value if no relationship exists between the number of host lineages and number of sites. P values were <0.001 for all of the networks (Supplementary Fig. 10).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw sequencing data and metadata are deposited in at DDBJ/EMBL/GenBank (BioProject PRJNA514023: SRA BioSamples SAMN10718335–SAMN10718821; Sanger Targeted Locus Study project accession numbers KCRE01000001–KCRE01010802). All of the sequencing data, metadata and other types of data used in this study are publicly available at figshare[69].

Received: 16 February 2019; Accepted: 2 August 2019;
Published online: 23 September 2019

References

1. Frelich, L. E. *Boreal Biome* (Oxford Bibliographies Online Datasets, 2013); https://doi.org/10.1093/obo/9780199830060-0085
2. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320**, 1444–1449 (2008).
3. Peng, C. et al. A drought-induced pervasive increase in tree mortality across Canada's boreal forests. *Nat. Clim. Change* **1**, 467–471 (2011).
4. Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A. Z. & Schepaschenko, D. G. Boreal forest health and global change. *Science* **349**, 819–822 (2015).
5. Gower, S. T. et al. Net primary production and carbon allocation patterns of boreal forest ecosystems. *Ecol. Appl.* **11**, 1395–1411 (2001).
6. Price, D. T. et al. Anticipating the consequences of climate change for Canada's boreal forest ecosystems. *Environ. Rev.* **21**, 322–365 (2013).
7. Lau, J. A. & Lennon, J. T. Rapid responses of soil microorganisms improve plant fitness in novel environments. *Proc. Natl Acad. Sci. USA* **109**, 14058–14062 (2012).
8. Martin, F. M., Uroz, S. & Barker, D. G. Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* **356**, eaad4501 (2017).
9. Clemmensen, K. E. et al. Roots and associated fungi drive long-term carbon sequestration in boreal forest. *Science* **339**, 1615–1618 (2013).
10. Treseder, K. K., Mack, M. C. & Cross, A. Relationships among fires, fungi, and soil dynamics in Alaskan boreal forests. *Ecol. Appl.* **14**, 1826–1838 (2004).
11. Arnold, A. E. et al. A phylogenetic estimation of trophic transition networks for ascomycetous fungi: are lichens cradles of symbiotrophic fungal diversification? *Syst. Biol.* **58**, 283–297 (2009).
12. Arnold, A. E. et al. Fungal endophytes limit pathogen damage in a tropical tree. *Proc. Natl Acad. Sci. USA* **100**, 15649–15654 (2003).
13. Busby, P. E. et al. Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biol.* **15**, e2001793 (2017).
14. Müller, D. B., Vogel, C., Bai, Y. & Vorholt, J. A. The plant microbiota: systems-level insights and perspectives. *Annu. Rev. Genet.* **50**, 211–234 (2016).
15. Rodriguez, R. J. et al. Stress tolerance in plants via habitat-adapted symbiosis. *ISME J.* **2**, 404–416 (2008).
16. Lutzoni, F. et al. Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat. Commun.* **9**, 5451 (2018).
17. Arnold, A. E. & Lutzoni, F. Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? *Ecology* **88**, 541–549 (2007).
18. U'Ren, J. M., Lutzoni, F., Miadlikowska, J., Laetsch, A. D. & Arnold, A. E. Host and geographic structure of endophytic and endolichenic fungi at a continental scale. *Am. J. Bot.* **99**, 898–914 (2012).
19. Zimmerman, N. B. & Vitousek, P. M. Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proc. Natl Acad. Sci. USA* **109**, 13022–13027 (2012).
20. van der Linde, S. et al. Environment and host as large-scale controls of ectomycorrhizal fungi. *Nature* **558**, 243–248 (2018).
21. Schoch, C. L. et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc. Natl Acad. Sci. USA* **109**, 6241–6246 (2012).
22. Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
23. Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
24. Soininen, J., McDonald, R. & Hillebrand, H. The distance decay of similarity in ecological communities. *Ecography* **30**, 3–12 (2007).
25. Yeoh, Y. K. et al. Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Nat. Commun.* **8**, 215 (2017).

26. Feurdean, A. et al. Tree migration-rates: narrowing the gap between inferred post-glacial rates and projected rates. *PLoS ONE* **8**, e71797 (2013).
27. Poulin, R., Krasnov, B. R., Mouillot, D. & Thieltges, D. W. The comparative ecology and biogeography of parasites. *Proc. R. Soc. B* **366**, 2379–2390 (2011).
28. Salgado-Salazar, C., Rossman, A. Y. & Chaverri, P. Not as ubiquitous as we thought: taxonomic crypsis, hidden diversity and cryptic speciation in the cosmopolitan fungus *Thelonectria discophora* (Nectriaceae, Hypocreales, Ascomycota). *PLoS ONE* **8**, e76737 (2013).
29. Golan, J. J. & Pringle, A. in *The Fungal Kingdom* (eds Heitman, J. et al.) 309–333 (ASM Press, 2017); https://doi.org/10.1128/microbiolspec.FUNK-0047-2016
30. Carbone, I. et al. T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads and metadata visualization: an example with the Pezizomycotina tree of life. *Bioinformatics* **33**, 1160–1168 (2017).
31. Giauque, H. & Hawkes, C. V. Climate affects symbiotic fungal endophyte diversity and performance. *Am. J. Bot.* **100**, 1435–1444 (2013).
32. Treseder, K. K., Marusenko, Y., Romero-Olivares, A. L. & Maltz, M. R. Experimental warming alters potential function of the fungal community in boreal forest. *Glob. Change Biol.* **22**, 3395–3404 (2016).
33. U'Ren, J. M. et al. Tissue storage and primer selection influence pyrosequencing-based inferences of diversity and community composition of endolichenic and endophytic fungi. *Mol. Ecol. Resour.* **14**, 1032–1048 (2014).
34. U'Ren, J. M. DNA extraction from fungal mycelium using Extract-n-Amp. *protocols.io* https://doi.org/10.17504/protocols.io.ga4bsgw (2016).
35. Higgins, K. L., Coley, P. D., Kursar, T. A. & Arnold, A. E. Culturing and direct PCR suggest prevalent host generalism among diverse fungal endophytes of tropical forest grasses. *Mycologia* **103**, 247–260 (2011).
36. U'Ren, J. M. & Arnold, A. E. DNA extraction protocol for plant and lichen tissues stored in CTAB. *protocols.io* https://doi.org/10.17504/protocols.io.fs8bnhw (2017).
37. U'Ren, J. M. & Arnold, A. E. Illumina MiSeq dual-barcoded two-step PCR amplicon sequencing protocol. *protocols.io* https://doi.org/10.17504/protocols.io.fs9bnh6 (2017).
38. Gardes, M. & Bruns, T. D. ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. *Mol. Ecol.* **2**, 113–118 (1993).
39. White, T. J. et al. in *PCR Protocols: A Guide to Methods and Applications* (eds Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J.) 315–322 (New York Academic Press, 1990).
40. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
41. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **31**, 3476–3482 (2015).
42. Bengtsson-Palme, J. et al. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* **4**, 914–919 (2013).
43. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
44. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
45. Abarenkov, K. et al. The UNITE database for molecular identification of fungi—recent updates and future perspectives. *New Phytol.* **186**, 281–285 (2010).
46. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
47. Huson, D. H. & Mitra, S. Introduction to the analysis of environmental sequences: metagenomics with MEGAN. *Methods Mol. Biol.* **856**, 415–429 (2012).
48. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
49. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
50. Edgar, R. C. UNCROSS2: identification of cross-talk in 16S rRNA OTU tables. Preprint at *bioRxiv* https://doi.org/10.1101/400762 (2018).
51. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
52. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/081257 (2016).
53. Berger, S. A. & Stamatakis, A. Aligning short reads to reference alignments and trees. *Bioinformatics* **27**, 2068–2075 (2011).
54. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
55. U'Ren, J. M. et al. Contributions of North American endophytes to the phylogeny, ecology, and taxonomy of Xylariaceae (Sordariomycetes, Ascomycota). *Mol. Phylogenet. Evol.* **98**, 210–232 (2016).
56. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1659 (2000).
57. Oksanen, J. et al. vegan: Community ecology package. R package version 2.4–6 (2018); https://CRAN.R-project.org/package=vegan
58. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2017).
59. Weiss, S. J. et al. Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data. Preprint at https://doi.org/10.7287/peerj.preprints.1157v1 (2015).
60. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
61. Legendre, P. & Anderson, M. J. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* **69**, 1–24 (1999).
62. McArdle, B. H. & Anderson, M. J. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82**, 290–297 (2001).
63. Borcard, D. & Legendre, P. Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. *Ecology* **93**, 1473–1481 (2012).
64. Nychka, D., Furrer, R., Paige, J. & Sain, S. Fields: tools for spatial data. R package version 9.8-3 (2015); https://CRAN.R-project.org/package=fields
65. Borcard, D. & Legendre, P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol. Model.* **153**, 51–68 (2002).
66. Dray, S., Legendre, P. & Peres-Neto, P. R. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Model.* **196**, 483–493 (2006).
67. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
68. Csardi, G. & Nepusz, T. igraph: Network analysis and visualization. R package version 0.7 (2014).
69. U'Ren, J. M. et al. Host availability drives distributions of fungal endophytes in the imperiled boreal realm. *figshare* https://doi.org/10.6084/m9.figshare.c.4327772 (2019).

Acknowledgements

We thank S. Irwin, L. Taylor, J. Stenlid, R. Andronova, A. Knorre, A. Dutbyeva, M. Zhurbenko, K. Arendt, E. Lefèvre, B. Ball, V. Wong, R. Oono, T. Gleason, J. Gonzales III, J. Riddle and K.-H. Chen for field and laboratory assistance; G. Hestmark, B. Hodkinson, S. LaGreca, J. Lendemer, B. McCune, L. Myllys, S. Stenroos and C. Truong for lichen identifications; B. Hurwitz, R. Steidl and G. Burleigh for helpful discussions; K. Youens-Clark and T. O'Connor for computational assistance; staff at the University of Arizona Genetics Core and D. New and A. Gerritsen at the University of Idaho IBEST Genomics Core for technical assistance; and M. Miller for deploying tools and databases used in T-BAS on CIPRES. This study was funded by the US National Science Foundation (NSF) Dimensions of Biodiversity program (A.E.A., DEB-1045766; I.C., DEB-1046167; G.M., DEB-1045608; F.L., DEB-1046065) and the Huron Mountain Wildlife Foundation (A.E.A.). N.B.Z. was supported by the Gordon and Betty Moore Foundation through grant number GBMF 2550.03 to the Life Sciences Research Foundation. The CIPRES RESTful API is supported by the National Institutes of Health (NIH; 5 R01 GM1264635), NSF (DBI-1759844) and an award (TG-DEB090011) of computer time and development support from the XSEDE project (also sponsored by NSF). Data collection that was performed by the IBEST Genomics Resources Core at the University of Idaho was supported in part by NIH (COBRE grant P30GM103324).

Author contributions

F.L., J.M. and A.E.A. conceived the study with I.C., G.M. and J.M.U. F.L., J.M., A.E.A., J.M.U. and G.M. conducted the fieldwork. J.M.U., J.M., F.L. and A.E.A. collected the data. J.M.U., A.E.A., N.B.Z., I.C. and F.L. developed the analyses. J.M.U., A.E.A. and N.B.Z. analysed the data. J.M.U., A.E.A. and F.L. wrote the paper with comments from all of the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-019-0975-2.

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

Correspondence and requests for materials should be addressed to A.E.A.

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

nature research

Corresponding author(s): A. Elizabeth Arnold

Last updated by author(s): Jul 23, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	USEARCH V5.2.32, QIIME v. 1.8, and ITSx were used to analyze Illumina sequence data. Initial editing of Sanger sequences was done in Mesquite with phred/phrap, followed by manual editing of sequence quality in Sequencher v. 5. MEGAN v. 5.11.3 was used to process BLASTn output and assign taxonomy. The Tree-Based Alignment Selector Toolkit (T-BAS) v. 2.1 (https://tbas.hpc.ncsu.edu/) was used to place ITS nrDNA-partial LSU nrDNA sequences in a phylogenetic context. Statistical analyses were done in R version 3.3.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequence data and metadata are deposited in at DDBJ/EMBL/GenBank (BioProject PRJNA514023: SRA BioSamples SAMN10718335- SAMN10718821; Sanger Targeted Locus Study project accession KCRE00000000). All sequence data, metadata, other data types, and code used in this study are publicly available in Figshare (ref 69).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	In this study we used culturing and culture-independent, next-generation sequencing (NGS) to examine fungal endophyte communities in phylogenetically diverse plant and lichen hosts (Magnoliophyta, Pinophyta, Monilophyta, Lycopodiophyta, Bryophyta, and lichens that comprised fungal mycobionts with Cyanobacteria, Chlorophyta, or both photobionts) that we collected in seven sites in North America and Eurasia that together circumscribe the global boreal belt.
Research sample	Research samples consisted of fresh, photosynthetic tissues of 498 plants and lichens from across the global boreal belt. The rationale for sampling was to ensure the greatest phylogenetic and ecological diversity of hosts within each sampling site. All endophyte data were generated by us. Data sources for soil fungi are cited in the main text and Methods.
Sampling strategy	We collected 498 individual plant- and lichen host collections in seven sites in North America and Eurasia that together circumscribe the global boreal belt. In each site we collected a minimum of 60 host individuals (at least 10 species of plants and thalli of at least 10 species of lichens (defined by mycobiont) in each of three replicate microsites). We previously used this sampling stategy to capture a statistically complete sample of culturable endophytic richness from a diversity of hosts within each site (i.e., such that estimated richness falls within the 95% confidence intervals of observed richness), thus permiting meaningful comparisons of diversity and composition among communities (U'Ren et al., 2012, Am. J. Bot.). Our sampling strategy included multiple spatial replicates per site and when possible, we included multiple taxonomic replicates per host lineage in each site (e.g., multiple species of Magnoliophyta, Pinophyta, Bryophyta, etc.). Sites were chosen to represent the breadth of mean annual precipitation (MAP) and mean annual temperature (MAT) encompassed by the boreal forest belt, while still allowing for similar plant and lichen species to be collected across the broadest geographical area to disentangle phylogenetic and spatial factors.
Data collection	Plant and lichen tissues from each host collection were surface-sterilized and processed following our previous methods (U'Ren et al., 2012, Am. J. Bot.). Under sterile conditions, 96 segments were chosen haphazardly for endophyte isolation on 2% malt extra agar, and an equal number were chosen haphazardly and placed in CTAB buffer for culture-independent analysis following U'Ren et al., (2016; Mol. Ecol. Res.). Portable laminar flow hoods facilitated sterile processing at remote locations, and sterile methods were used for all tissue processing steps. All emergent cultures were subject to DNA extraction, PCR amplification of the ITS nrDNA-partial LSU nrDNA locus, and bi-directional Sanger sequencing at the University of Arizona Genetics Core Facility. See Methods for details on processing and curation of Sanger sequences and fungal cultures. DNA extracted directly from host tissues was subject to amplification of the ITS nrDNA locus and sequenced with an Illumina MiSeq at the IBEST Genomics Resources Core at the University of Idaho. Demultiplexing was performed on raw data at the sequencing center. See Methods for detailed information on bioinformatic processing. Field collection information (lat/long, slope, canopy coverage, host species) was recorded at the time of sampling. When necessary, plant and lichen identifications were verified by taxonomic experts after collection.
Timing and spatial scale	Field collections from seven sites across North America (4 sites) and Eurasia (3 sites) were conducted at the height of the growing season (June-August) from 2011 to 2013. Details are provided in Supplementary Table 1. In each field site, plant and lichen samples were collected within a 5-hour time period and samples were processed for culturing and NGS within 48 hours after collection (with the exception of eastern Russia where materials were processed within 72 hours). Geographic distances between sites ranged from local (< 1-100 m) to global scales (up to 8,676 km).
Data exclusions	Only a small numbers of cultures or NGS sequences were excluded from analyses. Details are outlined in the Methods, Supplementary Tables, and Figure Legends. We have aimed for maximum transparency throughout.
Reproducibility	Although this was an observational study we verified our findings by repeated sampling in a focal site in Alaska (see details in Methods and Extended Data) and we verified our results using both culturing and next-generation sequencing methods.
Randomization	Our study was collections-based rather than experimental, thus our statistical analyses were mostly correlative. For analyses involving samples grouped by site or host lineage appropriate statistical tests were used to account for covariates.
Blinding	Host identity was not evident to researchers handling cultures or NGS samples, such that no systematic biases could be introduced inadvertently.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Details for each sampling site are provided in the Methods and Supplemental Table 1 (including forest composition, latitude, longitude, altitude, climate). Climate data for each site was obtained from the WorldClim database. In each site, we used tree cores, interviews with forestry agents, forestry data, and observations of fire damage (charcoal, scarring, and related indicators) to determine whether sites had experienced recent fires.

Location	Details for all sampling locations are provided in Supplementary Table 1 with further details on specific host locations provided in Supplementary Tables 2-3.
Access and import/export	All samples were imported by Arnold with USDA permission (APHIS PPQ permit # P526P-10-02180) and were housed according to permit regulations at the University of Arizona in approved quarantine space. All work was done with local collaborators or facilitators (Russia, eastern Russia, Sweden, Alaska, Québec) or with permission of granting agency (Michigan) or Parcs Canada (Wood Buffalo National Park, Alberta, Canada).
Disturbance	Plant and lichen hosts in each site were accessed via established roads and trails to minimize disturbances to the enviroment. We collected small amounts of material from each host to limit disturbance to the environment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging