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

Jana M. U’Ren¹, François Lutzoni², Jolanta Miadlikowska², Naupaka B. Zimmerman³, Ignazio Carbone⁴, Georgiana May⁵, A. Elizabeth Arnold⁶*

Affiliations: ¹Department of Biosystems Engineering and BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA. ²Department of Biology, Duke University, Durham, NC 27708, USA. ³Department of Biology, University of San Francisco, San Francisco, CA 94117, USA. ⁴Center for Integrated Fungal Research, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695, USA. ⁵Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108, USA. ⁶School of Plant Sciences and Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA.
Boreal forests represent the world’s largest terrestrial biome and provide ecosystem services of global importance. Highly imperiled by climate change, these forests host earth’s greatest phylogenetic diversity of endophytes, a hyperdiverse group of symbionts defined by their occurrence within living, symptomless plant and lichen tissues. Endophytes shape the ecological and evolutionary trajectories of plants and thus are key to the function and resilience of terrestrial ecosystems. A critical step in linking ecological functions of endophytes with those of their hosts is to understand their distributions at a global scale, but 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 unprecedented 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 geographic or environmental distance, is the main determinant of endophyte community composition and diversity. 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, endophyte distributions are linked directly to the availability of compatible hosts, highlighting the role of biotic interactions in shaping fungal communities across large spatial scales, and the threat of climate change to alter biological diversity and function in the imperiled boreal realm.

As the world’s largest terrestrial biome, boreal forests span > 11% of Earth’s land area and comprise ca. 30% of global forest cover. Boreal forests exert the greatest biogeophysical effects on mean global temperature and harbor a disproportionately high amount of carbon in soil, which – when combined with boreal vegetation – equals ca. 50% of the planet’s
atmospheric carbon. By 2100, warming due to climate change is expected to have a profound effect on biodiversity and species composition in boreal forests, yielding massive downstream effects on the net carbon balance and climate feedbacks driven by these high-latitude ecosystems.

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

The majority of fungal endophytes are transmitted horizontally, and over broad spatial scales their distributions generally reflect abiotic factors such as climate or geographic distance similar to free-living fungi in soil. However, host communities shift in composition with geography and climate, often confounding inferences about symbiont biogeography. 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 do so because of their broad consistency in vegetation types, climate, and phylogenetic composition of plant and lichen communities across continental to intercontinental scales.

We examined endophyte communities via culture-based sampling and culture-independent, next-generation sequencing (NGS) for 498 individual plant-and lichen host collections newly obtained in seven sites in North America and Eurasia that together circumscribe the global boreal belt (Fig. 1a, Supplementary Table 1). In 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/moss, rock, bark, or dead wood (Fig. 1b, Supplementary Fig. 1, Supplementary Tables 2 and 3). Overall, our sampling included at least 60 plant and lichen individual collections per site (range: 60-105) and an average of 19 host genera per site (range: 17-23) (Fig. 1). Geographic distances between individual host 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 mm² fragments for culturing and NGS. In total, we prepared > 46,000 fragments to isolate endophytes and an equivalent quantity for NGS (see Supplementary Tables 2 and 3). We obtained 11,975 endophyte isolates in culture and used the Sanger platform to sequence the fungal ITS nrDNA barcode locus and the adjacent, phylogenetically informative LSU nrDNA region for each isolate (Fig. 1c, Supplementary Tables 2 and 3). NGS analysis of host tissues generated > 5.8 million quality-filtered ITS2 nrDNA sequences of endophytes (Fig. 1c, Supplementary Tables 2 and 3).

Results and Discussion
Combined, the culture-based and NGS data sets included > 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 ca. 15-fold 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 a ca. fivefold increase in richness relative to culturing (NGS: 1,466.5 ± 21.3 OTUs; culturing: 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; Supplementary Methods). Sampling was sufficient for ecological inference (Supplementary Fig. 4) and repeated sampling at a focal site after three years 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 data set with > 44,000 OTUs observed in global surveys of soil fungi, including fungi from boreal soils, revealed that boreal endophytes were strikingly distinct. Only 1.5% of OTUs observed here were found in the global soil dataset (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 only to cultures, as in previous studies; 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 geographic 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\textsuperscript{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 (see Supplementary Table 5; Supplementary Fig. 6).

At the circumboreal scale, we predicted that dissimilarity of endophyte assemblages would correlate positively with geographic distance, consistent with distance decay\textsuperscript{24}. However, dissimilarity of endophyte assemblages could not be explained by geographic distance (Fig. 2a and B). Instead, host effects persisted at the global scale (Fig. 2c and d), reflecting the positive correlation between community dissimilarity of endophytes and genetic distance between host taxa across the circumboreal belt (Mantel test: $r = 0.20$, $P < 0.0001$; see also \textsuperscript{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, 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 artifacts of host distributions\textsuperscript{26}, rather than turnover due to inter-site distance per se. Accordingly, the slopes of species-area relationships for boreal endophytes are steep regardless of geographic scale (Supplementary Fig. 8, 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 geographic ranges due to the widespread availability of suitable partners\textsuperscript{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 to 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, 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 to j). The remaining OTUs appear to be host-generalists with wider geographic distributions than the more locally restricted specialists (Supplementary Fig. 10). While it is possible that apparent generalists contain cryptic species with more narrow distributions, 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 closely related to plant pathogens with transoceanic dispersal), such that dispersal limitation alone cannot explain their limited distributions. The availability of suitable hosts likely limits the geographic 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. The endophyte OTUs we isolated in culture were a representative subset of abundant OTUs obtained by NGS from the same host material (Supplementary Methods), but unlike the short sequences obtained by NGS were represented by longer sequencing reads.
containing regions that are informative for phylogenetic placement. 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 (Ascomycota) in both the culture-based and culture-independent data sets (i.e., Ascomycota; Fig. 4, Supplementary Figs. 12-16). We observed relatively wide host generalism and broad geographic distributions of endophyte-dominated clades in the 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 geographic 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 thus are fundamentally important to the dynamics and resilience of plant communities under climate stress. Experimental studies reveal direct sensitivity of boreal endophytes to warming and suggest altered functional roles with climate change. Our results suggest 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 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 detrimental to their continued persistence in the increasingly imperiled boreal realm.

Methods

Field collections

We collected fresh, photosynthetic tissues of diverse plants and lichens in seven sites across
North America and Eurasia (Fig. 1, Supplementary Table 1). Climate data were obtained from
the WorldClim database (www.worldclim.org) at 30 arcsecond resolution. There was no
evidence of recent fire in any focal site (based on tree cores, interviews with forestry agents,
forestry data, and observations of fire damage (charcoal, scarring, and related indicators)). Field
collections were conducted at the height of the growing season from 2011 to 2013
(Supplementary Table 1). In each site we collected fresh, mature, asymptomatic tissues of at
least 10 species of plants and thalli of at least 10 species of lichens (defined by mycobiont) in
each of three replicate microsites following \( ^{18} \) (Supplementary Tables 2 and 3; see also Fig. 1).
For each host we collected 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 tissue
processing steps described below.

**Endophyte isolation, DNA extraction, amplification, and Sanger sequencing**

Fresh tissues from each host collection were cut into 2 mm\(^2\) segments, which were surface-
sterilized following \( ^{18} \). Ninety-six segments were chosen haphazardly for endophyte isolation,
and an equal number were chosen haphazardly for culture-independent analysis (below)\( ^{33} \).
Endophytes were isolated on 2% malt extract agar (MEA) 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 internal
transcribed spacers and 5.8S gene (ITS nrDNA) and an adjacent portion of the nuclear ribosomal
large subunit (LSU nrDNA; ca. 500 base pairs; bp) was PCR-amplified as a single fragment (Fig.
NGS of endophyte communities: DNA extraction, amplification, and Illumina sequencing

Concurrently with culturing (above), we placed 96 surface-sterilized segments per host collection in CTAB buffer (1 M Tris HCl pH 8, 5 M NaCl, 0.5 M EDTA and 20 g CTAB) under sterile conditions. Tubes were stored at -80 °C until DNA was extracted. We extracted total genomic DNA with the MoBio PowerPlant Pro DNA Isolation Kit (Qiagen, Germantown, MD) and amplified and sequenced the fungal ITS nrDNA locus for each sample via a dual-barcoded, two-step library preparation process with the primer pair ITS1F/ITS4. We carried out PCR for each sample in three replicates. Amplification was verified on 2% agarose gels stained with SYBR Green I (Molecular Probes, Invitrogen, Carlsbad, CA, USA). Final PCR products were quantified fluorometrically with SYBR, normalized, and pooled in equimolar amounts. The final amplicon pool was purified with Agencourt AMPure XP beads following the manufacturer’s instructions (Beckman Coulter, Indianapolis, IN, USA). A BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) was used to determine DNA concentration and fragment size distribution of the final library prior to paired-end sequencing on an Illumina MiSeq with the Reagent Kit v3 (2x300 bp).

Bioinformatics and quality control

Raw Illumina data were demultiplexed and sequences representing PhiX and a “diversity shotgun library” (i.e., genomic DNA representing a non-fungal organism that is spiked into the run to improve cluster density during sequencing; IBEST Genomics Core, pers. comm.), as well
as sequences containing > 1 mismatches to the barcode and > 4 mismatches to primers, were removed. The remaining 9,942,458 reads corresponding to the ITS2 nrDNA region were trimmed for quality using a cutoff length of 170 bp and a maximum error rate of 1.0 in USEARCH v8.1.1861\textsuperscript{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 1.0.7\textsuperscript{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 bases downstream of the conserved region at the start of LSU nrDNA (i.e., 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 (i.e., singleton or doubleton OTUs) were removed\textsuperscript{33,40}.

**OTU clustering and taxonomic assignments**

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

A representative sequence from each OTU (chosen to represent the most abundant
sequence in the cluster) was queried first with ITSx\textsuperscript{42} to identify and subsequently remove OTUs lacking the ITS2 region. Sequences from the remaining OTUs were queried against NCBI nr (but excluding all environmental sequences) with BLAST\textit{n}\textsuperscript{46}. BLAST output was analyzed in MEGAN v. 5.11.3\textsuperscript{47} with default parameters for lowest common ancestor (LCA). OTUs representing lichen-forming fungi (i.e., the primary mycobiont, see Supplementary Table 3) or plant hosts, sequences with no hits, and/or sequences not classified to Fungi were removed from subsequent analyses. The remaining OTUs were queried against the UNITE fungal database\textsuperscript{45} with the RDP Classifier\textsuperscript{48} for taxonomic classification with a cutoff threshold of 80% confidence as implemented in QIIME v. 1.8\textsuperscript{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 (see \textsuperscript{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 denoised and clustered into zero-radius OTUs (i.e., zOTUs; analogous to amplicon sequence variants\textsuperscript{51}) with the UNOISE2 algorithm\textsuperscript{52} in USEARCH (see 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 also were clustered into OTUs independent of NGS reads following methods in \textsuperscript{18}. We observed a significant correlation in species richness per host when we compared 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 designated using full-length ITS
nrDNA-partial LSU nrDNA Sanger sequences were used for analyses of richness based on
culturing (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 at 99% ITS nrDNA sequence identity (to account for
differences due to different sequencing and bioinformatic methods between studies) with
UCLUST. Percent 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 (i.e., 3% of boreal endophyte OTUs were observed in the soil survey).

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 (see ), which were isolated, sequenced, and analyzed with the methods
described here (n = 1,042 cultures; 352 OTUs). Inter-site distances between boreal sites and
these sites ranged from 4,452 to > 10,000 km. Percent 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 of 524 OTUs), as well as all 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 with the Tree-Based Alignment Selector Toolkit (T-BAS) v. 2.1 (https://tbas.hpc.ncsu.edu/) with the evolutionary placement algorithm in RAxML for 10,805 cultures of boreal endophytes for which ITS nrDNA-partial LSU nrDNA sequences were obtained (Fig. 4; Supplementary Figs. 12-16). The reference Pezizomycotina tree in T-BAS is based on six loci. Settings used to place endophyte cultures within the reference Ascomycota tree with 5.8S nrDNA and partial LSU nrDNA sequences were: 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 (grey letters in Fig. 4) for RAxML analysis with 1000 bootstrap replicates following realignment in MAFFT, with data retained for all cultures (Fig. 4, Supplementary Figs. 12-16). Haplotype network analyses for sequences of Daldinia loculata (see for phylogenetic placement) were performed in T-BAS with TCS v. 1.21 (Supplementary Fig. 11).

**Statistical analyses**

**Richness and community structure**

We used an analysis of variance (ANOVA) 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 following (Fig. 2, Supplementary Fig. 6). We examined the relationship between endophyte species richness and environmental variables (mean annual temperature: MAT; mean annual precipitation: MAP), host lineage, and site with 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). Calculation of OTU richness estimates and rarefaction analyses were done with the vegan\textsuperscript{57} package in R\textsuperscript{58}.

We used NMDS ordinations based on Hellinger dissimilarity to visualize fungal community structure within each site (Supplementary Fig. 1). We used all host collections from our main sites (Fig. 1, with details in 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 $\geq 4$ sites: \textit{Rhododendron} (Magnoliophyta), \textit{Picea} (Pinophyta), \textit{Equisetum} (Monilophyta), \textit{Lycopodium} (Lycopodiophyta), \textit{Pleurozium} (Bryophyta), \textit{Cladonia} (chlorolichen), and \textit{Peltigera} (cyanolichens and tripartite lichens) (Fig. 2d). Read counts among samples differed by greater than 2-3x; 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 recommendations by \textsuperscript{59}. Due 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 OTU with $> 100$ reads. Data for three collections for each host species per site were combined to allow the NMDS analysis to converge (see Fig. 2d).

We used 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 (see also Supplementary Fig. 1), or as a function of host identity (i.e., 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
as described by 60–62. To account for the significant effect of host on endophyte community
structure, analyses at a 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 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 via 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, 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 (see 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 colonization18. We used t-tests to
compare isolation frequency between sampling years for plants and lichens separately
Because sampling intensity was 2x greater in the second sampling year we rarefied reads 1000x to compare richness (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 per above (Supplementary Fig. 5).

Spatial autocorrelation and distance decay

We computed Mantel correlograms of Hellinger community distance and intersite geographic distances (Fig. 2b) to quantify spatial autocorrelation. Intersite distances were measured with the Haversine method in the R package fields. 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 (dbRDA) constrained by principal components of neighbor matrices (PCNM), implemented in vegan as the “capscale” function. 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 Hellinger distance and Bray-Curtis dissimilarity in vegan to assess the importance of site-specific factors on endophyte community composition in focal host genera. If geographic distance affects endophyte community
composition, endophyte communities within a single host genus should cluster according to inter-site distances (see Fig. 1b, dendrogram at top). 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 sequence data representing the ribulose bisphosphate carboxylase large chain (rbcL) for plants. A lack of data for many mycobionts (i.e., ca. 66% of mycobionts for locus RPB1) precluded a similar analysis for lichens. The distance matrix of host rbcL sequences was computed from pairwise distances in mothur with default parameters. A Mantel test was implemented with vegan in R with the Pearson correlation method and 999 permutations.

**Species area relationships**

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

**Endophyte host associations**
We quantified and visualized the distribution of OTUs among major host lineages with networks constructed with the R package igraph 0.7.1. Networks were constructed for OTUs in each site (using endophytes from all host taxa; Supplementary Tables 2 and 3). Networks constructed at a circumglobal scale were restricted to (1) communities from a subset of 10 plant genera and five lichen genera, each of which was sampled in at least four sites (see Supplementary Fig. 9) or (2) endophyte communities from a representative genus for each major host lineage (see Fig. 3). We used Chi-square 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 (i.e., one site, two sites, etc.). Likelihood ratio tests were used to compute the probability of obtaining, by chance alone, a Chi-square value greater than the observed value if no relationship exists between the number of host lineages and number of sites. Probability values were < 0.001 for all networks (Supplementary Fig. 10).

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 numbers KCRE01000001-KCRE01010802). All sequence data, metadata, other data types, and code used in this study are publicly available in figshare (see 69).

References

   doi:10.1093/obo/9780199830060-0085

2. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of


13. Busby, P. E. et al. Research priorities for harnessing plant microbiomes in sustainable


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.


34. U’Ren, J. M. DNA extraction from fungal mycelium using Extract-n-Amp. (2016).
   doi:10.17504/protocols.io.ga4bsgw


43. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. 


65. Borcard, D. & Legendre, P. All-scale spatial analysis of ecological data by means of


https://doi.org/10.6084/m9.figshare.c.4327772.

### 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 discussion; K. Youens-Clark and T. O’Connor for computational assistance; 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 National Science Foundation Dimensions of Biodiversity program (AEA: DEB-1045766; IC: DEB-1046167; GM: DEB-1045608; FL: DEB-1046065) and the
Huron Mountain Wildlife Foundation (AEA). NBZ was supported by the Gordon and Betty Moore Foundation through Grant GBMF 2550.03 to the Life Sciences Research Foundation. The CIPRES RESTful API is supported by National Institutes of Health (NIH) 5 R01 GM1264635, NSF DBI-1759844, and a generous award (TG-DEB090011) of computer time and development support from the XSEDE project (also sponsored by the U.S. National Science Foundation). Data collection performed by the IBEST Genomics Resources Core at the University of Idaho was supported in part by NIH COBRE grant P30GM103324.

Author contributions Conceived of study: FL, JM, AEA, with IC, GM, and JMU. Conducted fieldwork: FL, JM, AEA, JMU, GM. Collected data: JMU, JM, FL, AEA. Developed analyses: JMU, AEA, NBZ, IC, FL. Analyzed data: JMU, AEA, NBZ. Wrote paper: JMU, AEA, FL, with comments from all authors.

Author information The authors declare no competing interests. Correspondence and requests for materials should be addressed to AEA (arnold@ag.arizona.edu).

Figure Legends

Fig. 1. Geographic location, climate, and host information for 498 individual host collections sampled for endophytes in seven boreal sites. a, Sampling sites. Map source: Base map © Mapbox © OpenStreetMap, see https://www.mapbox.com/about/maps/ and https://www.openstreetmap.org/copyright. b, Number of host individuals and host genera collected in each site, depicted with relative geographic distances among sites (top dendrogram)
and relationships of photobiont host lineages (left). Lichen photobionts include Chlorophyta or Cyanobacteria, which can occur alone within thalli (i.e., in chlorolichens† or cyanolichens‡, respectively) or together in one thallus (i.e., tripartite lichen*, counted above only once as Chlorophyta because the tripartite lichen thalli we collected were dominated (in area or volume) by the algal photobionts). c, Fungal barcode locus: nuclear ribosomal internal transcribed spacers and 5.8S gene, sequenced for cultures with a portion of the nuclear ribosomal large subunit (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, Mantel correlogram as a function of geographic distance classes among seven sites illustrates a lack of geographic autocorrelation. c, A quantile box plot illustrates variation in richness among host lineages (ANOVA with post-hoc Tukey’s HSD (letters)). d, NMDS for one host genus per lineage (colors) at a circumboreal scale reveals differences in endophyte communities among host genera, including those host genera sampled in ≥ 4 sites (shapes): Rhododendron (Magnoliophyta), Picea (Pinophyta), Equisetum (Monilophyta), Lycopodium (Lycopodiophyta), Pleurozium (Bryophyta), Cladonia (chlorolichen), and Peltigera (cyanolichens and tripartite lichens). Statistics reflect host x site interaction in PERMANOVA after combining data for three replicate individuals for each host species per site (see Methods for more detail and Supplementary Table 6).

Fig. 3. Networks reveal host affiliations of endophyte OTUs at local and circumboreal scales. Nodes represent OTUs. Edges connect OTUs to host lineage(s) in which they were found.
a-g, Networks by site, with node diameter proportional to $\log_{10}$ read abundance. Color indicates the number of host lineages in which an OTU was observed. OTU richness and read depth shown for each host lineage/site. h-j, Networks for circumboreal data set, with 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 $\geq 4$ sites. Asterisks (*) indicate cyanolichens.

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 endophytes isolates, obtained by culturing, with rings of metadata (host, site, continent) in color. Reference taxa are shown with colored 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. Diameter of each circular tree represents relative abundance of each focal class. Lichens have photobionts as described in Fig. 1. Support values are shown in Supplementary Figs. 12-16.
Fig. 1.

[Image of a map showing mean annual precipitation and temperature across different regions.]

Fig. 2.

[Image of a diagram showing host lineage distribution across different sites.]
Fig. 3.
Fig. 4