

1 **Boreal forests represent the world’s largest terrestrial biome and provide ecosystem**
2 **services of global importance. Highly imperiled by climate change, these forests host**
3 **earth’s greatest phylogenetic diversity of endophytes, a hyperdiverse group of symbionts**
4 **defined by their occurrence within living, symptomless plant and lichen tissues.**
5 **Endophytes shape the ecological and evolutionary trajectories of plants and thus are key to**
6 **the function and resilience of terrestrial ecosystems. A critical step in linking ecological**
7 **functions of endophytes with those of their hosts is to understand their distributions at a**
8 **global scale, but turnover in host taxa with geography and climate can confound insights**
9 **into endophyte biogeography. As a result, global drivers of endophyte diversity and**
10 **distributions are not known. Here, we leverage unprecedented sampling from**
11 **phylogenetically diverse boreal plants and lichens across North America and Eurasia to**
12 **show that host filtering in distinctive environments, rather than turnover with geographic**
13 **or environmental distance, is the main determinant of endophyte community composition**
14 **and diversity. We reveal the distinctiveness of boreal endophytes relative to soil fungi**
15 **worldwide and endophytes from diverse temperate biomes, highlighting a high degree of**
16 **global endemism. Overall, endophyte distributions are linked directly to the availability of**
17 **compatible hosts, highlighting the role of biotic interactions in shaping fungal communities**
18 **across large spatial scales, and the threat of climate change to alter biological diversity and**
19 **function in the imperiled boreal realm.**

20 As the world’s largest terrestrial biome, boreal forests span > 11% of Earth’s land area
21 and comprise ca. 30% of global forest cover¹. Boreal forests exert the greatest biogeophysical
22 effects on mean global temperature and harbor a disproportionately high amount of carbon in
23 soil, which – when combined with boreal vegetation – equals ca. 50% of the planet’s

1 atmospheric carbon². By 2100, warming due to climate change is expected to have a profound
2 effect on biodiversity and species composition in boreal forests³, yielding massive downstream
3 effects on the net carbon balance and climate feedbacks driven by these high-latitude
4 ecosystems⁴⁻⁶.

5 Plant-associated microbial communities are increasingly recognized for their potential to
6 facilitate rapid acclimation of plants to novel stressors, especially within threatened biomes^{7,8}.
7 Soilborne and root-associated fungi are critical to nutrient cycling, soil dynamics, and ecosystem
8 productivity and resilience in boreal ecosystems^{9,10}. Long under-studied because of their cryptic
9 occurrence in healthy above-ground tissues, fungal endophytes that occur within photosynthetic
10 tissues of plants and in association with photosynthetic partners in lichens¹¹ also are key players
11 in host health, productivity, and stress mitigation¹²⁻¹⁵. Endophytes originated contemporaneously
12 with the origin of land plants¹⁶ and comparative studies reveal that they reach their greatest
13 phylogenetic diversity in boreal forests, exceeding that even of tropical regions¹⁷. In highly
14 imperiled boreal forests, understanding the distributions of endophytes is a critical first step in
15 linking their ecological functions with those of their hosts, and key to interpreting the resilience
16 of ecosystems such as forests to environmental change².

17 The majority of fungal endophytes are transmitted horizontally, and over broad spatial
18 scales their distributions generally reflect abiotic factors such as climate or geographic
19 distance¹⁸⁻¹⁹ similar to free-living fungi in soil²⁰. However, host communities shift in
20 composition with geography and climate, often confounding inferences about symbiont
21 biogeography¹⁸. As a result, there is a need to disentangle deterministic processes such as host-
22 and environmental filtering from neutral processes such as dispersal and drift as drivers of
23 endophyte diversity and distributions at a circumglobal scale. Boreal forests represent a unique

1 opportunity to do so because of their broad consistency in vegetation types, climate, and
2 phylogenetic composition of plant and lichen communities across continental to intercontinental
3 scales¹.

4 We examined endophyte communities via culture-based sampling and culture-
5 independent, next-generation sequencing (NGS) for 498 individual plant- and lichen host
6 collections newly obtained in seven sites in North America and Eurasia that together
7 circumscribe the global boreal belt (Fig. 1a, Supplementary Table 1). In each site we collected
8 photosynthetic tissues from living, asymptomatic plants representing Magnoliophyta, Pinophyta,
9 Monilophyta, Lycopodiophyta, and Bryophyta, as well as lichens that comprised fungal
10 mycobionts with Cyanobacteria, Chlorophyta, or both photobionts on soil/moss, rock, bark, or
11 dead wood (Fig. 1b, Supplementary Fig. 1, Supplementary Tables 2 and 3). Overall, our
12 sampling included at least 60 plant and lichen individual collections per site (range: 60-105) and
13 an average of 19 host genera per site (range: 17-23) (Fig. 1). Geographic distances between
14 individual host collections ranged from local (< 1-100 m) to global scales (up to 8,676 km) (Fig.
15 1a). Host tissues were surface-sterilized and cut into 2 mm² fragments¹⁸ for culturing and NGS.
16 In total, we prepared > 46,000 fragments to isolate endophytes and an equivalent quantity for
17 NGS (see Supplementary Tables 2 and 3). We obtained 11,975 endophyte isolates in culture and
18 used the Sanger platform to sequence the fungal ITS nrDNA barcode locus²¹ and the adjacent,
19 phylogenetically informative LSU nrDNA region for each isolate (Fig. 1c, Supplementary Tables
20 2 and 3). NGS analysis of host tissues generated > 5.8 million quality-filtered ITS2 nrDNA
21 sequences of endophytes (Fig. 1c, Supplementary Tables 2 and 3).

22

23 **Results and Discussion**

1 Combined, the culture-based and NGS data sets included > 6,000 operational taxonomic units
2 (OTUs) in five fungal phyla, including a minimum of seven classes of Ascomycota and seven
3 classes of Basidiomycota (Supplementary Fig. 2). Endophyte richness values inferred by
4 culturing and NGS were correlated positively, independent of host lineage or sequencing depth
5 (Supplementary Fig. 3). Overall, richness based on NGS was ca. 15-fold greater than that
6 inferred by culturing from the equivalent quantity of host tissue (Supplementary Tables 2 and 3).
7 When NGS data were subsampled to match the number of sequenced cultures, NGS provided a
8 ca. fivefold increase in richness relative to culturing (NGS: $1,466.5 \pm 21.3$ OTUs; culturing:
9 315.0 ± 7.0 OTUs). Culturing and NGS recovered the same classes and orders of Ascomycota,
10 albeit in different proportions, whereas NGS recovered a higher diversity of Basidiomycota
11 (Supplementary Fig. 2; Supplementary Methods). Sampling was sufficient for ecological
12 inference (Supplementary Fig. 4) and repeated sampling at a focal site after three years showed
13 that a single sampling event was representative of the local endophyte community over the
14 timescale of our study (Supplementary Fig. 5).

15 Comparison of the entire data set with > 44,000 OTUs observed in global surveys of soil
16 fungi²², including fungi from boreal soils, revealed that boreal endophytes were strikingly
17 distinct. Only 1.5% of OTUs observed here were found in the global soil dataset²²
18 (Supplementary Table 4). Similarly, only 2.5% of OTUs observed here were found in
19 comparable surveys of endophytes from the temperate zone (12% when data were restricted only
20 to cultures, as in previous studies¹⁸; Supplementary Table 4).

21 These findings underscore the tremendous richness of boreal endophytes and the distinct
22 niche they occupy as symbionts. As such, we evaluated the importance of host identity, climate,
23 and geographic distance in structuring endophyte assemblages at local to circumboreal scales.

1 Within each site, host identity was the major predictor of endophyte community structure
2 (Supplementary Fig. 1). Host genus explained an average of 58% of the variation in endophyte
3 community composition within sites (51-68%; Supplementary Fig. 1). As for soil fungi^{22,23},
4 endophyte richness was correlated positively with mean annual precipitation (MAP)
5 (Supplementary Table 5). For boreal endophytes, however, host lineage had greater explanatory
6 power than MAP in our models (see Supplementary Table 5; Supplementary Fig. 6).

7 At the circumboreal scale, we predicted that dissimilarity of endophyte assemblages
8 would correlate positively with geographic distance, consistent with distance decay²⁴. However,
9 dissimilarity of endophyte assemblages could not be explained by geographic distance (Fig. 2a
10 and B). Instead, host effects persisted at the global scale (Fig. 2c and d), reflecting the positive
11 correlation between community dissimilarity of endophytes and genetic distance between host
12 taxa across the circumboreal belt (Mantel test: $r = 0.20$, $P < 0.0001$; see also ²⁵ for similar
13 correlations for root-associated bacteria, but at local scales). These host effects were modulated
14 by site-specific factors, the importance of which varied among host lineages (Supplementary Fig.
15 7, Supplementary Table 6). Thus, assemblages of boreal endophytes appear to largely reflect
16 biotic filtering by hosts in the context of distinctive environments, microclimates, or historical
17 artifacts of host distributions²⁶, rather than turnover due to inter-site distance per se. Accordingly,
18 the slopes of species-area relationships for boreal endophytes are steep regardless of geographic
19 scale (Supplementary Fig. 8, Supplementary Table 7).

20 For horizontally transmitted symbionts, host colonization requires both dispersal to the
21 host and symbiotic establishment. Endophyte OTUs with wide host ranges might be predicted to
22 have large geographic ranges due to the widespread availability of suitable partners²⁷. To test this
23 prediction, we used networks to visualize the associations of endophyte OTUs with hosts at local

1 and circumboreal scales (Fig. 3). Even when analyses were restricted to the most common
2 OTUs, an average of 64% of OTUs were affiliated with members of only one host lineage in
3 each site (Fig. 3a to g). The number of host lineages in which an OTU was found was a poor
4 predictor of OTU abundance, suggesting that OTUs were not designated inappropriately as
5 specialists simply because they were rare (Supplementary Fig. 9, Supplementary Methods).
6 When scaled to the circumboreal level, an average of 24% of the most common OTUs still
7 associated with members of only one host lineage (Fig. 3h to j). The remaining OTUs appear to
8 be host-generalists with wider geographic distributions than the more locally restricted
9 specialists (Supplementary Fig. 10). While it is possible that apparent generalists contain cryptic
10 species with more narrow distributions²⁸, haplotype analysis of sequences representing the most
11 widespread generalist OTU reveals a global distribution of the most abundant amplicon sequence
12 variants (ASVs) (Supplementary Fig. 11). Geographically restricted and specialist OTUs
13 represent diverse genera with different spore sizes and discharge methods (including endophytes
14 closely related to plant pathogens with transoceanic dispersal²⁹), such that dispersal limitation
15 alone cannot explain their limited distributions. The availability of suitable hosts likely limits the
16 geographic distributions of specialists and drives the high global richness of endophytes at a
17 circumboreal scale.

18 Acknowledging the evolutionary relatedness among OTUs provides an important
19 framework for understanding ecological patterns. At present, the relatively short sequences
20 generated by NGS for fungi usually cannot be placed reliably in community-scale phylogenetic
21 analyses^{21,30}. The endophyte OTUs we isolated in culture were a representative subset of
22 abundant OTUs obtained by NGS from the same host material (Supplementary Methods), but
23 unlike the short sequences obtained by NGS were represented by longer sequencing reads

1 containing regions that are informative for phylogenetic placement³⁰. By placing these cultured
2 endophytes in a robust phylogenetic framework for the first time, we detected distinctive
3 evolutionary trajectories in each focal class of the most prevalent phylum (Ascomycota) in both
4 the culture-based and culture-independent data sets (i.e., Ascomycota; Fig. 4, Supplementary
5 Figs. 12-16). We observed relatively wide host generalism and broad geographic distributions of
6 endophyte-dominated clades in the Sordariomycetes and Pezizomycetes, which affiliate
7 especially frequently with lichens (Supplementary Fig. 2). In contrast, endophytes in classes such
8 as Dothideomycetes and Leotiomycetes often had narrower host- and geographic distributions,
9 and were observed more frequently in plants (Supplementary Fig. 2).

10 Fungal endophytes influence the functional traits, ecological dynamics, and evolutionary
11 trajectories of their hosts, and thus are fundamentally important to the dynamics and resilience of
12 plant communities under climate stress¹⁴. Experimental studies reveal direct sensitivity of boreal
13 endophytes to warming and suggest altered functional roles with climate change^{31,32}. Our results
14 suggest endophytes of boreal plants and lichens are distinctive, hyperdiverse, and distributed in a
15 manner that reflects the presence of compatible hosts at local to circumglobal scales. Thus, shifts
16 in climate that lead to local and regional extirpation of plants and lichens⁴ are likely to result in
17 the rapid loss of endophyte diversity locally. As a consequence, boreal plant and lichen
18 communities globally may face a loss of symbiont-conferred resilience—a change detrimental to
19 their continued persistence in the increasingly imperiled boreal realm.

20

21 **Methods**

22 **Field collections**

23 We collected fresh, photosynthetic tissues of diverse plants and lichens in seven sites across

1 North America and Eurasia (Fig. 1, Supplementary Table 1). Climate data were obtained from
2 the WorldClim database (www.worldclim.org) at 30 arcsecond resolution. There was no
3 evidence of recent fire in any focal site (based on tree cores, interviews with forestry agents,
4 forestry data, and observations of fire damage (charcoal, scarring, and related indicators)). Field
5 collections were conducted at the height of the growing season from 2011 to 2013
6 (Supplementary Table 1). In each site we collected fresh, mature, asymptomatic tissues of at
7 least 10 species of plants and thalli of at least 10 species of lichens (defined by mycobiont) in
8 each of three replicate microsites following ¹⁸ (Supplementary Tables 2 and 3; see also Fig. 1).
9 For each host we collected random subset of photosynthetic tissues that, for long-lived
10 individuals or tissues, encompassed multiple years of growth. Portable laminar flow hoods
11 facilitated sterile processing at remote locations, and sterile methods were used for all tissue
12 processing steps described below.

13

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

15 Fresh tissues from each host collection were cut into 2 mm² segments, which were surface-
16 sterilized following ¹⁸. Ninety-six segments were chosen haphazardly for endophyte isolation,
17 and an equal number were chosen haphazardly for culture-independent analysis (below)³³.
18 Endophytes were isolated on 2% malt extract agar (MEA) under sterile conditions¹⁸. Fungi that
19 emerged from tissue pieces were vouchered in sterile water and deposited at the Robert L.
20 Gilbertson Mycological Herbarium at the University of Arizona (Supplementary Table 8). We
21 extracted total genomic DNA directly from each fungal isolate³⁴. The nuclear ribosomal internal
22 transcribed spacers and 5.8S gene (ITS nrDNA) and an adjacent portion of the nuclear ribosomal
23 large subunit (LSU nrDNA; ca. 500 base pairs; bp) was PCR-amplified as a single fragment (Fig.

1 1c), sequenced bidirectionally, and processed manually¹⁸. High quality Sanger sequence data
2 were obtained for 10,805 of 11,975 isolates (Supplementary Tables 2, 3, and 8).

3

4 **NGS of endophyte communities: DNA extraction, amplification, and Illumina sequencing**

5 Concurrently with culturing (above), we placed 96 surface-sterilized segments per host collection
6 in CTAB buffer (1 M Tris HCl pH 8, 5 M NaCl, 0.5 M EDTA and 20 g CTAB)^{33,35} under sterile
7 conditions. Tubes were stored at -80 °C until DNA was extracted. We extracted total genomic
8 DNA with the MoBio PowerPlant Pro DNA Isolation Kit (Qiagen, Germantown, MD)³⁶ and
9 amplified and sequenced the fungal ITS nrDNA locus for each sample via a dual-barcoded, two-
10 step library preparation process³⁷ with the primer pair ITS1F/ITS4^{38,39}. We carried out PCR for
11 each sample in three replicates. Amplification was verified on 2% agarose gels stained with
12 SYBR Green I (Molecular Probes, Invitrogen, Carlsbad, CA, USA). Final PCR products were
13 quantified fluorometrically with SYBR, normalized, and pooled in equimolar amounts. The final
14 amplicon pool was purified with Agencourt AMPure XP beads following the manufacturer's
15 instructions (Beckman Coulter, Indianapolis, IN, USA). A BioAnalyzer 2100 (Agilent
16 Technologies, Santa Clara, CA, USA) was used to determine DNA concentration and fragment
17 size distribution of the final library prior to paired-end sequencing on an Illumina MiSeq with the
18 Reagent Kit v3 (2x300 bp).

19

20 **Bioinformatics and quality control**

21 Raw Illumina data were demultiplexed and sequences representing PhiX and a “diversity
22 shotgun library” (i.e., genomic DNA representing a non-fungal organism that is spiked into the
23 run to improve cluster density during sequencing; IBEST Genomics Core, pers. comm.), as well

1 as sequences containing > 1 mismatches to the barcode and > 4 mismatches to primers, were
2 removed. The remaining 9,942,458 reads corresponding to the ITS2 nrDNA region were
3 trimmed for quality using a cutoff length of 170 bp and a maximum error rate of 1.0 in
4 USEARCH v8.1.1861^{40,41}, resulting in 4,553,953 high-quality sequences. To combine Sanger
5 sequences from cultures with Illumina sequences for direct comparisons, we first used ITSx
6 1.0.7⁴² to identify Sanger sequences that did not contain at least 50 bp of either ITS1 or ITS2
7 nrDNA. These sequences (n = 86) were removed. For the remaining 10,719 Sanger sequences
8 (Supplementary Tables 2 and 3), all bases downstream of the conserved region at the start of
9 LSU nrDNA (i.e., 3' end) were removed and the 5' end of the sequences were trimmed to a
10 length of 170 bp to match the exact length and start position of Illumina sequences. Sanger and
11 Illumina sequences were dereplicated in parallel and OTUs represented by only one or two
12 Illumina sequences (i.e., singleton or doubleton OTUs) were removed^{33,40}.

14 **OTU clustering and taxonomic assignments**

15 After these filtering steps, dereplicated sequences from both the culture-based and NGS analyses
16 were clustered into operational taxonomic units (OTUs) at 95% sequence similarity with
17 UPARSE-OTU algorithm⁴³ as implemented in USEARCH⁴⁰, a decision based on the clustering
18 results of the mock community (see Supplementary Methods). In addition to *de novo* chimera
19 checking performed during clustering⁴⁴, representative sequences for each OTU were subjected
20 to reference-based chimera checking using the UNITE⁴⁵ database with UCHIME⁴⁴. Raw
21 Illumina reads and all Sanger reads were mapped back to chimera-checked OTUs to construct an
22 OTU table containing > 6 million reads and > 6,200 OTUs.

23 A representative sequence from each OTU (chosen to represent the most abundant

1 sequence in the cluster) was queried first with ITSx⁴² to identify and subsequently remove OTUs
2 lacking the ITS2 region. Sequences from the remaining OTUs were queried against NCBI nr (but
3 excluding all environmental sequences) with BLASTn⁴⁶. BLAST output was analyzed in
4 MEGAN v. 5.11.3⁴⁷ with default parameters for lowest common ancestor (LCA). OTUs
5 representing lichen-forming fungi (i.e., the primary mycobiont, see Supplementary Table 3) or
6 plant hosts, sequences with no hits, and/or sequences not classified to Fungi were removed from
7 subsequent analyses. The remaining OTUs were queried against the UNITE fungal database⁴⁵
8 with the RDP Classifier⁴⁸ for taxonomic classification with a cutoff threshold of 80% confidence
9 as implemented in QIIME v. 1.8⁴⁹ (Supplementary Fig. 2). Analyses of the phylogenetically
10 diverse mock community confirmed our bioinformatic methods for (1) removal of spurious
11 OTUs resulting from erroneous sequences; (2) low prevalence of tag-switching (see ⁵⁰) among
12 samples (< 1% of OTUs); and (3) correct estimates of species boundaries for phylogenetically
13 diverse taxa present in the mock community (Supplementary Table 9). Representative analyses
14 described below were repeated with data denoised and clustered into zero-radius OTUs (i.e.,
15 zOTUs; analogous to amplicon sequence variants⁵¹) with the UNOISE2 algorithm⁵² in
16 USEARCH (see Supplementary Methods). In this context rare taxa were more abundant, but our
17 main results did not differ appreciably.

18 Sanger sequences containing the entire ITS nrDNA-partial LSU nrDNA region also were
19 clustered into OTUs independent of NGS reads following methods in ¹⁸. We observed a
20 significant correlation in species richness per host when we compared OTU richness based on
21 full-length ITS nrDNA-partial LSU nrDNA Sanger sequences and OTU richness based on
22 trimmed ITS2 nrDNA reads (to match NGS read length, see above) (Pearson correlation: $r =$
23 0.98 , $P < 0.0001$; Supplementary Tables 2 and 3). OTUs designated using full-length ITS

1 nrDNA-partial LSU nrDNA Sanger sequences were used for analyses of richness based on
2 culturing (below; Supplementary Tables 2 and 3).

3

4 **Comparison of boreal endophytes to a global survey of fungi from soil**

5 A representative sequence for each OTU was clustered with representative sequences for 44,563
6 fungal OTUs from a global survey of soil²² at 99% ITS nrDNA sequence identity (to account for
7 differences due to different sequencing and bioinformatic methods between studies) with
8 UCLUST⁴⁰. Percent overlap was calculated as the number of boreal endophyte OTUs that
9 clustered with a fungal OTU from soil divided by the total number of OTUs from soil fungi
10 (Supplementary Table 4). Similar results were obtained when clustering was repeated using 97%
11 sequence similarity (i.e., 3% of boreal endophyte OTUs were observed in the soil survey²²).

12

13 **Comparisons of boreal endophytes to endophytes of temperate plants and lichens**

14 We compared the overlap of boreal endophytes with endophytes from plants and lichens in a
15 temperate semideciduous forest, temperate coniferous forest, and subtropical scrub forest of
16 North America (see ¹⁸), which were isolated, sequenced, and analyzed with the methods
17 described here (n = 1,042 cultures; 352 OTUs). Inter-site distances between boreal sites and
18 these sites ranged from 4,452 to > 10,000 km. Percent overlap was calculated as the number of
19 OTUs that contained both boreal and temperate endophytes divided by the total OTUs for
20 cultured endophytes only (12.0%; 63 of 524 OTUs), as well as all cultures plus NGS reads
21 (2.5%; 153 out of 6,152 OTUs) (Supplementary Table 4).

22

23 **Phylogenetic analyses**

1 Phylogenetic placement of endophytes in the Ascomycota was inferred with the Tree-Based
2 Alignment Selector Toolkit (T-BAS) v. 2.1 (<https://tbas.hpc.ncsu.edu/>)³⁰ with the evolutionary
3 placement algorithm in RAxML⁵³ for 10,805 cultures of boreal endophytes for which ITS
4 nrDNA-partial LSU nrDNA sequences were obtained (Fig. 4; Supplementary Figs. 12-16). The
5 reference Pezizomycotina tree in T-BAS is based on six loci³⁰. Settings used to place endophyte
6 cultures within the reference Ascomycota tree with 5.8S nrDNA and partial LSU nrDNA
7 sequences were: UNITE filter engaged, 1.0 sequence identity, genetic distance score = 10
8 standard deviations, likelihood weights (fast), with the outgroup selected. Each major class of
9 Pezizomycotina was then selected (grey letters in Fig. 4) for RAxML analysis with 1000
10 bootstrap replicates following realignment in MAFFT⁵⁴, with data retained for all cultures (Fig.
11 4, Supplementary Figs. 12-16). Haplotype network analyses for sequences of *Daldinia loculata*
12 (see ⁵⁵ for phylogenetic placement) were performed in T-BAS with TCS v. 1.21⁵⁶
13 (Supplementary Fig. 11).

14

15 **Statistical analyses**

16 **Richness and community structure**

17 We used an analysis of variance (ANOVA) to compare richness among sites and host lineages
18 for cultures and NGS after accounting for differences in sequencing depth. Richness was defined
19 for the analysis by calculating the residuals of OTU richness in relation to the square-root of the
20 number of reads following ²² (Fig. 2, Supplementary Fig. 6). We examined the relationship
21 between endophyte species richness and environmental variables (mean annual temperature:
22 MAT; mean annual precipitation: MAP), host lineage, and site with linear mixed models
23 (Supplementary Table 5). We compared total richness among host lineages and sites for both

1 cultures and NGS data using rarefaction (Supplementary Fig. 4). Calculation of OTU richness
2 estimates and rarefaction analyses were done with the *vegan*⁵⁷ package in R⁵⁸.

3 We used NMDS ordinations based on Hellinger dissimilarity to visualize fungal
4 community structure within each site (Supplementary Fig. 1). We used all host collections from
5 our main sites (Fig. 1, with details in Supplementary Tables 2 and 3). We used the same
6 approach for the analysis across seven sites that span the circumboreal belt; however, for each
7 major host lineage we used data from a single representative genus sampled in ≥ 4 sites:

8 *Rhododendron* (Magnoliophyta), *Picea* (Pinophyta), *Equisetum* (Monilophyta), *Lycopodium*
9 (Lycopodiophyta), *Pleurozium* (Bryophyta), *Cladonia* (chlorolichen), and *Peltigera*
10 (cyanolichens and tripartite lichens) (Fig. 2d). Read counts among samples differed by greater
11 than 2-3x; thus, to remove the effect of differential sequencing depth we rarefied the number of
12 NGS reads per host to the lowest number of sequences following recommendations by⁵⁹. Due to
13 the preponderance of zeros in the OTU matrix, non-convergence of the ordination search, and
14 high stress values, NMDS analyses at the circumglobal scale were restricted to OTU with > 100
15 reads. Data for three collections for each host species per site were combined to allow the NMDS
16 analysis to converge (see Fig. 2d).

17 We used PERMANOVA with the Hellinger distance metric to assess the significance of
18 community similarity as a function of host genus and lineage (plant/algal phylum or mycobiont
19 order) at the local scale (see also Supplementary Fig. 1), or as a function of host identity (i.e.,
20 lineage and genus), site, and/or environmental variables (MAT, MAP) at the circumglobal scale
21 (Supplementary Table 6). In these analyses, data from multiple microsites were not combined
22 (whereas in Fig. 2d, data from multiple microsites were combined to achieve convergence of the
23 NMDS analysis and lower stress values, see above). Site explained a greater proportion of

1 variation in endophyte community composition than MAP and/or MAT; therefore, we used site
2 as an explanatory variable because it encapsulates both climate as well as other site-specific
3 factors. PERMANOVA were implemented using the "adonis" function in the R library vegan⁵⁷
4 as described by ⁶⁰⁻⁶². To account for the significant effect of host on endophyte community
5 structure, analyses at a circumglobal scale were conducted with the entire dataset as well as
6 various subsets of hosts including (1) only plants; (2) only lichens; or (3) various combinations
7 of 16 plant and lichen genera, each found in a minimum of four sites (Supplementary Table 6).

8

9 **Relationship between richness estimates from cultures and NGS**

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

17

18 **Assessment of interannual variation in endophyte communities**

19 We compared the isolation frequency, richness, and community composition of endophytes
20 isolated in culture in one site (AKE, Supplementary Table 1) in summer of 2008 (see ¹⁸) and
21 2011 (Supplementary Fig. 5). Isolation frequency, defined as the percentage of tissue segments
22 containing cultivable fungi, was used as a proxy for host tissue colonization¹⁸. We used t-tests to
23 compare isolation frequency between sampling years for plants and lichens separately

1 (Supplementary Fig. 5). Because sampling intensity was 2x greater in the second sampling year
2 we rarefied reads 1000x to compare richness (Supplementary Fig. 5). We used PERMANOVA
3 to test for differences in endophyte community composition as a function of sampling year and
4 visualized endophyte communities with NMDS per above (Supplementary Fig. 5).

6 **Spatial autocorrelation and distance decay**

7 We computed Mantel correlograms of Hellinger community distance and intersite geographic
8 distances (Fig. 2b) to quantify spatial autocorrelation⁶³. Intersite distances were measured with
9 the Haversine method in the R package fields⁶⁴. Correlation coefficients were computed after
10 999 permutations. Relationships between community distance and intersite distances for Sanger
11 sequences and NGS data were plotted to visualize distance decay (Fig. 2a), and Mantel tests
12 were computed to test for a correlation. To test the significance of site and host lineage on
13 communities while constraining variation attributable to distance alone, we used distance-based
14 redundancy analysis (dbRDA) constrained by principal components of neighbor matrices
15 (PCNM), implemented in vegan as the “capscale” function^{61,65,66}. The "ordiR2step" function in
16 vegan was used for forward model choice solely on adjusted R² and P-values. RDA also was
17 used to assess variation attributable to spatial eigenvectors alone, after accounting for host
18 lineage and site effects.

20 **Hierarchical clustering of endophyte communities in focal host genera**

21 We used UPGMA average linkage clustering with Hellinger distance and Bray-Curtis
22 dissimilarity in vegan⁵⁷ to assess the importance of site-specific factors on endophyte community
23 composition in focal host genera. If geographic distance affects endophyte community

1 composition, endophyte communities within a single host genus should cluster according to
2 inter-site distances (see Fig. 1b, dendrogram at top). Instead, UPGMA dendrograms appear to
3 illustrate site-specific factors, the importance of which varied among host genera (Supplementary
4 Fig. 7).

6 **Relationship of host genetic distance and endophyte community dissimilarity**

7 We used Mantel tests to examine the correlation between host genetic distance and endophyte
8 community dissimilarity. Endophyte community dissimilarity was defined with Hellinger
9 distance (see above) and host genetic distance was estimated by analysis of sequence data
10 representing the ribulose biphosphate carboxylase large chain (*rbcL*) for plants. A lack of data
11 for many mycobionts (i.e., ca. 66% of mycobionts for locus *RPBI*) precluded a similar analysis
12 for lichens. The distance matrix of host *rbcL* sequences was computed from pairwise distances in
13 mothur⁶⁷ with default parameters. A Mantel test was implemented with *vegan*⁵⁷ in R with the
14 Pearson correlation method and 999 permutations.

16 **Species area relationships**

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

23 **Endophyte host associations**

1 We quantified and visualized the distribution of OTUs among major host lineages with networks
2 constructed with the R package igraph 0.7.1⁶⁸. Networks were constructed for OTUs in each site
3 (using endophytes from all host taxa; Supplementary Tables 2 and 3). Networks constructed at a
4 circumglobal scale were restricted to (1) communities from a subset of 10 plant genera and five
5 lichen genera, each of which was sampled in at least four sites (see Supplementary Fig. 9) or (2)
6 endophyte communities from a representative genus for each major host lineage (see Fig. 3). We
7 used Chi-square tests to evaluate the null hypothesis that the number of host lineages used by an
8 endophyte OTU was consistent regardless of the number of sites in which that OTU was found
9 (i.e., one site, two sites, etc.). Likelihood ratio tests were used to compute the probability of
10 obtaining, by chance alone, a Chi-square value greater than the observed value if no relationship
11 exists between the number of host lineages and number of sites. Probability values were < 0.001
12 for all networks (Supplementary Fig. 10).

13
14 **Data Availability.** Raw sequence data and metadata are deposited in at DDBJ/EMBL/GenBank
15 (BioProject PRJNA514023: SRA BioSamples SAMN10718335- SAMN10718821; Sanger
16 Targeted Locus Study project accession numbers KCRE01000001-KCRE01010802). All
17 sequence data, metadata, other data types, and code used in this study are publicly available in
18 figshare (see ⁶⁹).

20 **References**

- 21 1. Frelich, L. E. Boreal Biome. *Oxford Bibliographies Online Datasets* (2013).
22 doi:10.1093/obo/9780199830060-0085
- 23 2. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of

- 1 forests. *Science* **320**, 1444–1449 (2008).
- 2 3. Peng, C. *et al.* A drought-induced pervasive increase in tree mortality across Canada's
3 boreal forests. *Nat. Clim. Chang.* **1**, 467 (2011).
- 4 4. Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A. Z. & Schepaschenko, D. G.
5 Boreal forest health and global change. *Science* **349**, 819–822 (2015).
- 6 5. Gower, S. T. *et al.* Net primary production and carbon allocation patterns of boreal forest
7 ecosystems. *Ecol. Appl.* **11**, 1395–1411 (2001).
- 8 6. Price, D. T. *et al.* Anticipating the consequences of climate change for Canada's boreal
9 forest ecosystems. *Environ. Rev.* **21**, 322–365 (2013).
- 10 7. Lau, J. A. & Lennon, J. T. Rapid responses of soil microorganisms improve plant fitness in
11 novel environments. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14058–14062 (2012).
- 12 8. Martin, F. M., Uroz, S. & Barker, D. G. Ancestral alliances: Plant mutualistic symbioses
13 with fungi and bacteria. *Science* **356**, (2017).
- 14 9. Clemmensen, K. E. *et al.* Roots and associated fungi drive long-term carbon sequestration
15 in boreal forest. *Science* **339**, 1615–1618 (2013).
- 16 10. Treseder, K. K., Mack, M. C. & Cross, A. Relationships among fires, fungi, and soil
17 dynamics in Alaskan boreal forests. *Ecol. Appl.* **14**, (2004).
- 18 11. Arnold, A. E. *et al.* A phylogenetic estimation of trophic transition networks for
19 ascomycetous fungi: are lichens cradles of symbiotrophic fungal diversification? *Syst. Biol.*
20 **58**, 283–297 (2009).
- 21 12. Arnold, A. E. *et al.* Fungal endophytes limit pathogen damage in a tropical tree. *Proc. Natl.*
22 *Acad. Sci. U. S. A.* **100**, 15649–15654 (2003).
- 23 13. Busby, P. E. *et al.* Research priorities for harnessing plant microbiomes in sustainable

- 1 agriculture. *PLoS Biol.* **15**, e2001793 (2017).
- 2 14. Müller, D. B., Vogel, C., Bai, Y. & Vorholt, J. A. The plant microbiota: systems-level
3 insights and perspectives. *Annu. Rev. Genet.* **50**, 211–234 (2016).
- 4 15. Rodriguez, R. J. *et al.* Stress tolerance in plants via habitat-adapted symbiosis. *ISME J.* **2**,
5 404–416 (2008).
- 6 16. Lutzoni, F. *et al.* Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat.*
7 *Commun.* **9**, 5451 (2018).
- 8 17. Arnold, A. E. & Lutzoni, F. Diversity and host range of foliar fungal endophytes: are
9 tropical leaves biodiversity hotspots? *Ecology* **88**, 541–549 (2007).
- 10 18. U'Ren, J. M., Lutzoni, F., Miadlikowska, J., Laetsch, A. D. & Arnold, A. E. Host and
11 geographic structure of endophytic and endolichenic fungi at a continental scale. *Am. J. Bot.*
12 **99**, 898–914 (2012).
- 13 19. Zimmerman, N. B. & Vitousek, P. M. Fungal endophyte communities reflect environmental
14 structuring across a Hawaiian landscape. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13022–13027
15 (2012).
- 16 20. van der Linde, S. *et al.* Environment and host as large-scale controls of ectomycorrhizal
17 fungi. *Nature* **558**, 243–248 (2018).
- 18 21. Schoch, C. L. *et al.* Nuclear ribosomal internal transcribed spacer (ITS) region as a
19 universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 6241–6246
20 (2012).
- 21 22. Tedersoo, L. *et al.* Global diversity and geography of soil fungi. *Science* **346**, 1256688
22 (2014).
- 23 23. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**,

- 1 233–237 (2018).
- 2 24. Soininen, J., McDonald, R. & Hillebrand, H. The distance decay of similarity in ecological
3 communities. *Ecography* **30**, 3–12 (2007).
- 4 25. Yeoh, Y. K. *et al.* Evolutionary conservation of a core root microbiome across plant phyla
5 along a tropical soil chronosequence. *Nat. Commun.* **8**, 215 (2017).
- 6 26. Feurdean, A. *et al.* Tree migration-rates: narrowing the gap between inferred post-glacial
7 rates and projected rates. *PLoS One* **8**, e71797 (2013).
- 8 27. Poulin, R., Krasnov, B. R., Mouillot, D. & Thieltges, D. W. The comparative ecology and
9 biogeography of parasites. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2379–2390 (2011).
- 10 28. Salgado-Salazar, C., Rossman, A. Y. & Chaverri, P. Not as ubiquitous as we thought:
11 taxonomic crypsis, hidden diversity and cryptic speciation in the cosmopolitan fungus
12 *Theλονectria discophora* (Nectriaceae, Hypocreales, Ascomycota). *PLoS One* **8**, e76737
13 (2013).
- 14 29. Golan, J. J. & Pringle, A. Long-distance dispersal of Fungi. *Microbiol Spectr* **5**, (2017).
- 15 30. Carbone, I. *et al.* T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based
16 placement, alignment downloads and metadata visualization: an example with the
17 Pezizomycotina tree of life. *Bioinformatics* **33**, 1160–1168 (2017).
- 18 31. Giauque, H. & Hawkes, C. V. Climate affects symbiotic fungal endophyte diversity and
19 performance. *Am. J. Bot.* **100**, 1435–1444 (2013).
- 20 32. Treseder, K. K., Marusenko, Y., Romero-Olivares, A. L. & Maltz, M. R. Experimental
21 warming alters potential function of the fungal community in boreal forest. *Glob. Chang.*
22 *Biol.* **22**, 3395–3404 (2016).
- 23 33. U'Ren, J. M. *et al.* Tissue storage and primer selection influence pyrosequencing-based

- 1 inferences of diversity and community composition of endolichenic and endophytic fungi.
2 *Mol. Ecol. Resour.* **14**, 1032–1048 (2014).
- 3 34. U'Ren, J. M. DNA extraction from fungal mycelium using Extract-n-Amp. (2016).
4 doi:10.17504/protocols.io.ga4bsgw
- 5 35. Higgins, K. L., Coley, P. D., Kursar, T. A. & Arnold, A. E. Culturing and direct PCR
6 suggest prevalent host generalism among diverse fungal endophytes of tropical forest
7 grasses. *Mycologia* **103**, 247–260 (2011).
- 8 36. U'Ren, J. M. & Arnold, A. E. DNA extraction protocol for plant and lichen tissues stored in
9 CTAB. (2017). doi:10.17504/protocols.io.fs8bnhw
- 10 37. U'Ren, J. M. & Arnold, A. E. Illumina MiSeq Dual-barcoded Two-step PCR Amplicon
11 Sequencing Protocol. (2017). doi:10.17504/protocols.io.fs9bnh6
- 12 38. Gardes, M. & Bruns, T. D. ITS primers with enhanced specificity for basidiomycetes--
13 application to the identification of mycorrhizae and rusts. *Mol. Ecol.* **2**, 113–118 (1993).
- 14 39. White, T. J., Bruns, T., Lee, S., Taylor, J. W. & Others. Amplification and direct
15 sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to*
16 *methods and applications* **18**, 315–322 (1990).
- 17 40. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
18 **26**, 2460–2461 (2010).
- 19 41. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for next-
20 generation sequencing reads. *Bioinformatics* **31**, 3476–3482 (2015).
- 21 42. Bengtsson-Palme, J. *et al.* Improved software detection and extraction of ITS1 and ITS2
22 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental
23 sequencing data. *Methods Ecol. Evol.* **4**, 914–919 (2013).

- 1 43. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads.
2 *Nat. Methods* **10**, 996–998 (2013).
- 3 44. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves
4 sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
- 5 45. Abarenkov, K. *et al.* The UNITE database for molecular identification of fungi--recent
6 updates and future perspectives. *New Phytol.* **186**, 281–285 (2010).
- 7 46. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
8 search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- 9 47. Huson, D. H. & Mitra, S. Introduction to the analysis of environmental sequences:
10 metagenomics with MEGAN. *Methods Mol. Biol.* **856**, 415–429 (2012).
- 11 48. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid
12 assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.*
13 **73**, 5261–5267 (2007).
- 14 49. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing
15 data. *Nat. Methods* **7**, 335–336 (2010).
- 16 50. Edgar, R. C. UNCROSS2: identification of cross-talk in 16S rRNA OTU tables. *bioRxiv*
17 400762 (2018). doi:10.1101/400762
- 18 51. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon
19 data. *Nat. Methods* **13**, 581–583 (2016).
- 20 52. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon
21 sequencing. *bioRxiv* 081257 (2016). doi:10.1101/081257
- 22 53. Berger, S. A. & Stamatakis, A. Aligning short reads to reference alignments and trees.
23 *Bioinformatics* **27**, 2068–2075 (2011).

- 1 54. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
2 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- 3 55. U'Ren, J. M. *et al.* Contributions of North American endophytes to the phylogeny, ecology,
4 and taxonomy of Xylariaceae (Sordariomycetes, Ascomycota). *Mol. Phylogenet. Evol.* **98**,
5 210–232 (2016).
- 6 56. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene
7 genealogies. *Mol. Ecol.* **9**, 1657–1659 (2000).
- 8 57. Oksanen, J. *et al.* The vegan package. *Community ecology package* **10**, 631–637 (2007).
- 9 58. Team, R. C. R: A language and environment for statistical computing. Vienna, Austria: R
10 Foundation for Statistical Computing; 2017.
- 11 59. Weiss, S. J. *et al.* Effects of library size variance, sparsity, and compositionality on the
12 analysis of microbiome data. (PeerJ PrePrints, 2015). doi:10.7287/peerj.preprints.1157v1
- 13 60. Anderson, M. J. A new method for non-parametric multivariate analysis of variance.
14 *Austral Ecol.* **26**, 32–46 (2001).
- 15 61. Legendre, P. & Anderson, M. J. Distance-based redundancy analysis: testing multispecies
16 responses in multifactorial ecological experiments. *Ecol. Monogr.* **69**, 1–24 (1999).
- 17 62. McArdle, B. H. & Anderson, M. J. Fitting multivariate models to community data: a
18 comment on distance-based redundancy analysis. *Ecology* **82**, 290–297 (2001).
- 19 63. Borcard, D. & Legendre, P. Is the Mantel correlogram powerful enough to be useful in
20 ecological analysis? A simulation study. *Ecology* **93**, 1473–1481 (2012).
- 21 64. Nychka, D., Furrer, R., Paige, J. & Sain, S. Fields: tools for spatial data. doi: 10.5065.
22 (2015).
- 23 65. Borcard, D. & Legendre, P. All-scale spatial analysis of ecological data by means of

- 1 principal coordinates of neighbour matrices. *Ecol. Modell.* **153**, 51–68 (2002).
- 2 66. Dray, S., Legendre, P. & Peres-Neto, P. R. Spatial modelling: a comprehensive framework
3 for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Modell.* **196**, 483–
4 493 (2006).
- 5 67. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-
6 supported software for describing and comparing microbial communities. *Appl. Environ.*
7 *Microbiol.* **75**, 7537–7541 (2009).
- 8 68. Csardi, G. & Nepusz, T. igraph: Network analysis and visualization. *R package version 0. 7*
9 (2014).
- 10 69. U'Ren, J.M. *et al.* Data from: Host availability drives distributions of fungal endophytes in
11 the imperiled boreal realm. figshare. Collection. (2019)
12 <https://doi.org/10.6084/m9.figshare.c.4327772>.

13

14 **Acknowledgements** We thank S. Irwin, L. Taylor, J. Stenlid, R. Andronova, A. Knorre, A.
15 Dutbyeva, M. Zhurbenko, K. Arendt, E. Lefèvre, B. Ball, V. Wong, R. Oono, T. Gleason, J.
16 Gonzales III, J. Riddle, and K-H. Chen for field and laboratory assistance; G. Hestmark, B.
17 Hodgkinson, S. LaGreca, J. Lendemer, B. McCune, L. Myllys, S. Stenroos, and C. Truong for
18 lichen identifications; B. Hurwitz, R. Steidl, and G. Burleigh for helpful discussion; K. Youens-
19 Clark and T. O'Connor for computational assistance; the University of Arizona Genetics Core
20 and D. New and A. Gerritsen at the University of Idaho IBEST Genomics Core for technical
21 assistance; and M. Miller for deploying tools and databases used in T-BAS on CIPRES. This
22 study was funded by the National Science Foundation Dimensions of Biodiversity program
23 (AEA: DEB-1045766; IC: DEB-1046167; GM: DEB-1045608; FL: DEB-1046065) and the

1 Huron Mountain Wildlife Foundation (AEA). NBZ was supported by the Gordon and Betty
2 Moore Foundation through Grant GBMF 2550.03 to the Life Sciences Research Foundation. The
3 CIPRES RESTful API is supported by National Institutes of Health (NIH) 5 R01 GM1264635,
4 NSF DBI-1759844, and a generous award (TG-DEB090011) of computer time and development
5 support from the XSEDE project (also sponsored by the U.S. National Science Foundation). Data
6 collection performed by the IBEST Genomics Resources Core at the University of Idaho was
7 supported in part by NIH COBRE grant P30GM103324.

8

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

13

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

16

17 **Figure Legends**

18

19 **Fig. 1. Geographic location, climate, and host information for 498 individual host**
20 **collections sampled for endophytes in seven boreal sites. a**, Sampling sites. Map source: Base
21 map © Mapbox © OpenStreetMap, see <https://www.mapbox.com/about/maps/> and
22 <https://www.openstreetmap.org/copyright>. **b**, Number of host individuals and host genera
23 collected in each site, depicted with relative geographic distances among sites (top dendrogram)

1 and relationships of photobiont host lineages (left). Lichen photobionts include Chlorophyta or
2 Cyanobacteria, which can occur alone within thalli (i.e., in chlorolichens† or cyanolichens‡,
3 respectively) or together in one thallus (i.e., tripartite lichens*, counted above only once as
4 Chlorophyta because the tripartite lichen thalli we collected were dominated (in area or volume)
5 by the algal photobionts). **c**, Fungal barcode locus: nuclear ribosomal internal transcribed spacers
6 and 5.8S gene, sequenced for cultures with a portion of the nuclear ribosomal large subunit (ITS
7 nrDNA-partial LSU nrDNA). NGS data represent the ITS2 nrDNA region.

8

9 **Fig. 2. Host identity structures endophyte communities at a circumboreal scale. a**, We
10 observed no evidence of distance decay in endophyte community similarity. **b**, Mantel
11 correlogram as a function of geographic distance classes among seven sites illustrates a lack of
12 geographic autocorrelation. **c**, A quantile box plot illustrates variation in richness among host
13 lineages (ANOVA with post-hoc Tukey's HSD (letters)). **d**, NMDS for one host genus per
14 lineage (colors) at a circumboreal scale reveals differences in endophyte communities among
15 host genera, including those host genera sampled in ≥ 4 sites (shapes): *Rhododendron*
16 (Magnoliophyta), *Picea* (Pinophyta), *Equisetum* (Monilophyta), *Lycopodium* (Lycopodiophyta),
17 *Pleurozium* (Bryophyta), *Cladonia* (chlorolichen), and *Peltigera* (cyanolichens and tripartite
18 lichens). Statistics reflect host x site interaction in PERMANOVA after combining data for three
19 replicate individuals for each host species per site (see Methods for more detail and
20 Supplementary Table 6).

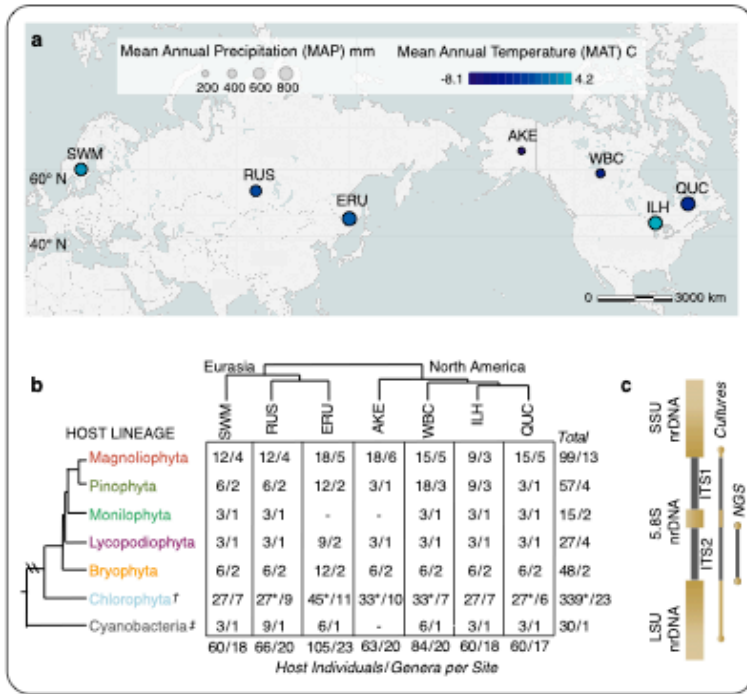
21

22 **Fig. 3. Networks reveal host affiliations of endophyte OTUs at local and circumboreal**
23 **scales.** Nodes represent OTUs. Edges connect OTUs to host lineage(s) in which they were found.

1 **a-g**, Networks by site, with node diameter proportional to \log_{10} read abundance. Color indicates
2 the number of host lineages in which an OTU was observed. OTU richness and read depth
3 shown for each host lineage/site. **h-j**, Networks for circumboreal data set, with node diameter is
4 proportional to the number of sites in which the OTU was observed. Host lineages are
5 represented by a single host genus sampled in ≥ 4 sites. Asterisks (*) indicate cyanolichens.
6

7 **Fig. 4. Evolutionary context of endophyte-host associations revealed by phylogenetic**
8 **analyses of the most species-rich fungal phylum (Ascomycota).** Phylogenetic placement of
9 endophytes was inferred in T-BAS³⁰. Trees show endophytes isolates, obtained by culturing,
10 with rings of metadata (host, site, continent) in color. Reference taxa are shown with colored
11 branches and no metadata³⁰. **a**, Pezizomycotina, the largest subphylum of Ascomycota, with
12 letters corresponding to panels **b-f**, which represent the most endophyte-rich classes of
13 Pezizomycotina. Diameter of each circular tree represents relative abundance of each focal class.
14 Lichens have photobionts as described in Fig. 1. Support values are shown in Supplementary
15 Figs. 12-16.

1 Fig 1.

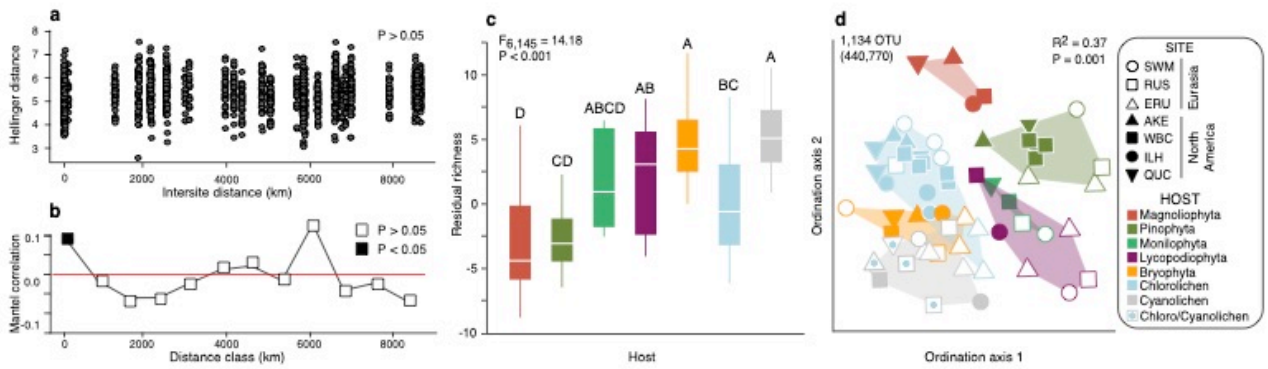


2

3

4

5 Fig 2.



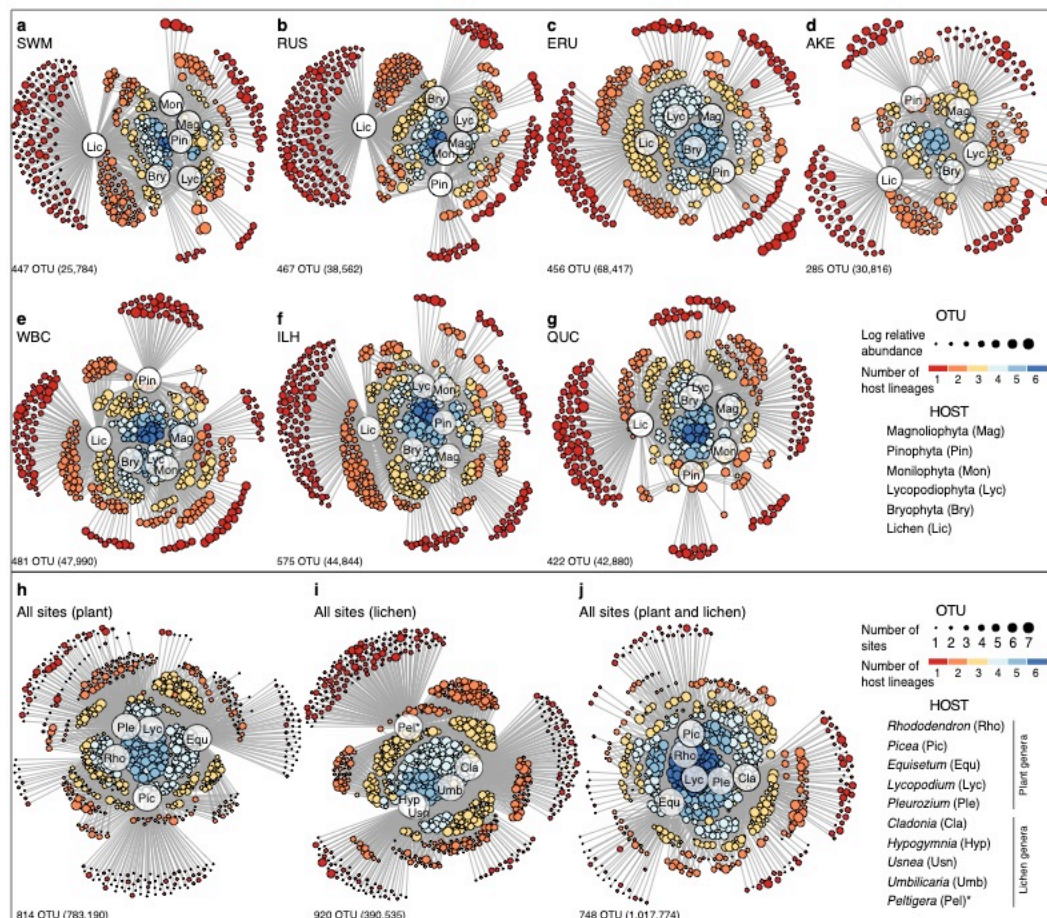
6

7

8

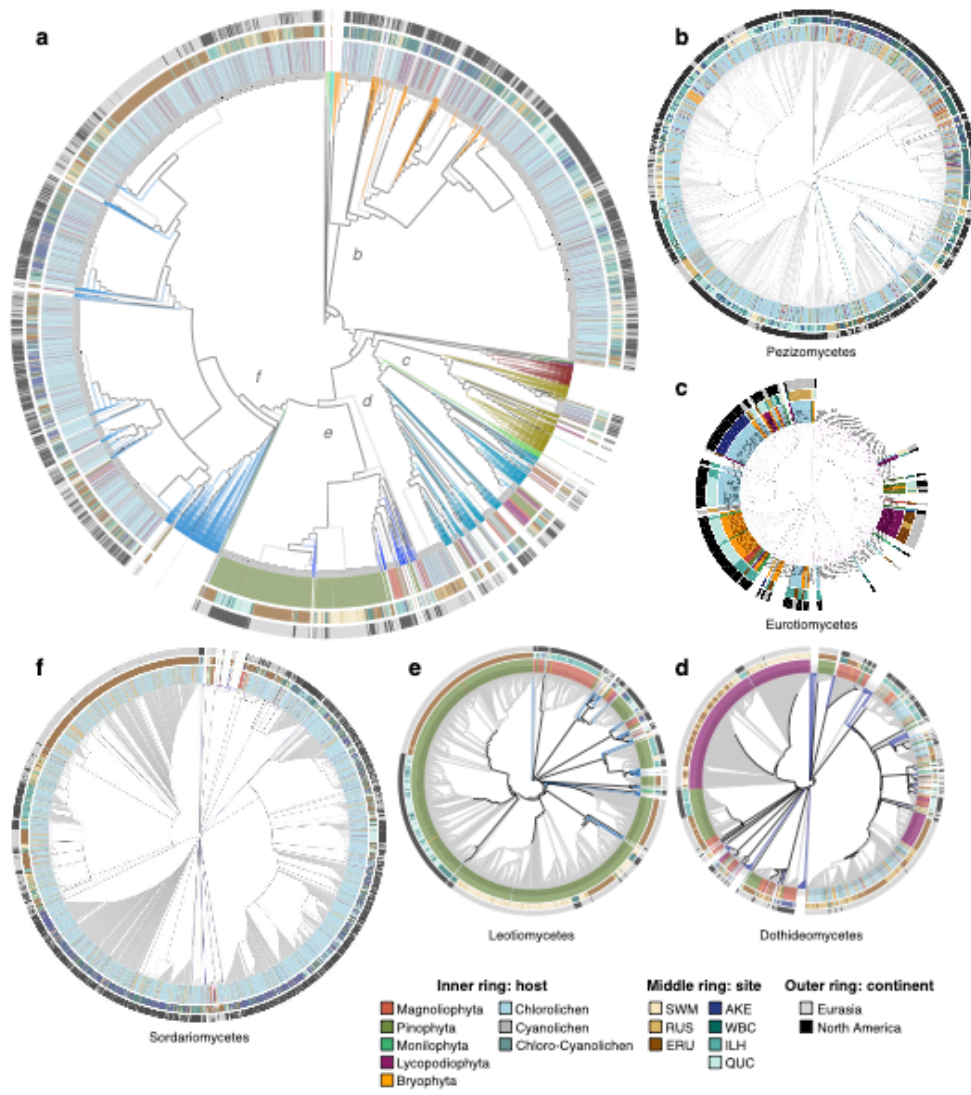
9

1 Fig. 3.



2
3
4
5
6
7
8
9
10

1 Fig. 4



2