nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-019-0975-2

In the format provided by the authors and unedited.

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

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

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

Supplementary information for

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

Jana M. U'Ren, François Lutzoni, Jolanta Miadlikowska, Naupaka B. Zimmerman, Ignazio Carbone, Georgiana May, A. Elizabeth Arnold

Contents: Supplementary Methods and Supplementary Figures 1-16

Supplementary Methods

Field collections

Sites were identified to represent the breadth of mean annual precipitation (MAP) and mean annual temperature (MAT) encompassed by the boreal forest belt[1]. Microsites in each site were located ca. 30 m apart along a 100 m transect. Each microsite was approximately 75 m^2 in area, typically extending laterally from the transect so that each microsite was discrete. In some sites we made additional collections to ensure representation of the phylogenetic and ecological diversity of hosts within sites (Supplementary Tables 1-3). Unless noted, each host shown in Supplementary Tables 2 and 3 was collected three times, resulting in three host collections per taxon per site (Fig. 1). Tissue samples were stored on ice or at 4 °C (plants) or naturally dried (lichens) and processed within 72 hours of collection except for those from ERU (samples from the 'main' site were processed within 96 hours of collection; see Supplementary Tables 2 and 3).

Bioinformatics analyses and quality control

Primer choice and rationale for sequencing

Fungal-specific primers that amplify the ITS2 region for fungi while excluding plants are not presently available[70]. We therefore amplified the entire ITS nrDNA region with the forward primer ITS1F (which results in the fewest reads for plants during *in silico* PCR[71]) and the reverse primer ITS4. Thus forward NGS reads (i.e., R1) corresponded to the ITS1 region and the reverse NGS reads (i.e., R2) corresponded to the ITS2 region. Analysis of R1 and R2 reads yielded similar results (data not shown). We analyzed the R2 reads for two key reasons. First, analysis of R2 allows comparisons to previous studies of soil fungi that used the ITS2 region (e.g., [22]). Second, because Sanger sequences for cultures extended into the LSU region (see Fig. 1) they could be trimmed to match the exact start and end positions of the NGS sequences, providing a basis for comparison of the OTUs generated by culture-based and NGS approaches. Sanger sequences were trimmed manually as part of the sequence validation and editing process (see methods in [18]) and thus did not all encompass the exact start position of R1 as generated by NGS.

Negative controls for NGS

We sequenced negative controls representing DNA extraction blanks and PCR negative controls.

Extraction blanks were generated for each MoBio PowerPro kit lot number used to extract DNA as well as intermittently throughout extractions. We used PCR negative template controls (NTC) with water as template for each 96-well plate of PCR1. NTCs from PCR1 were carried through to PCR2 to ensure no cross-contamination during PCR2 setup. In addition, a separate NTC using water as template was used for each 96-well plate of PCR2. We observed no bands on agarose gels indicating contamination. All PCR NTCs were pooled and sequenced with Illumina. We used these negative controls to assess the potential for OTUs in our dataset to represent laboratory contamination. In total, we removed 274,950 reads (18 OTUs) of potential contaminants found in negative controls (ca. 3.5% of total quality-controlled reads).

Positive controls for NGS

We sequenced a mock community that contained 32 phylogenetically diverse taxa representing four phyla (Chytridiomycota, Mucoromycota, Basidiomycota, and Ascomycota) as a positive control (Supplementary Table 9). DNA from each taxon was amplified individually in PCR1, quantified, and normalized to 1 ng/µl. Normalized PCR1 products for each taxon were pooled in equimolar amounts and used as the template for PCR2. All samples and positive and negative controls were sequenced in the same run, so that there was no concern about variation among sequencing runs.

We used the mock community data to validate our bioinformatic methods to accurately estimate species boundaries. Our previous assessment of four endophyte-rich genera in the Sordariomycetes and Dothideomycetes demonstrated that 5% ITS nrDNA divergence (i.e., 95% sequence similarity) conservatively estimated sister species boundaries when compared against published phylogenies[55,72]. Here, we clustered OTUs in USEARCH with UPARSE[40,43] at 95% and 97% ITS2 nrDNA sequence identity, and we compared the number of reads matching taxa in the mock for five replicates of the mock community. For all replicates, OTUs defined at 95% ITS2 nrDNA sequence similarity resulted in the most accurate estimates of species boundaries and richness. Representative sequences for OTUs were correctly assigned to each taxon with an average of 99.99% sequence similarity between the known sequence and sequences recovered from NGS. This approach also limited the number of spurious OTUs resulting from sequencing errors (i.e., each mock taxon was represented by a single OTU; see [33]). We also confirmed that

bioinformatic methods limited spurious OTUs due to barcode or tag shifting (i.e., 'cross-talk'[50]). After careful examination, we identified three OTUs in the mock dataset that likely resulted from tag switching (i.e., each OTU was represented by a single read in the mock, but numerous reads in real samples). Although this estimate (< 1%) is low, we used beta diversity indices that take abundance into account (i.e., Hellinger, Bray-Curtis), rather than using only presence/absence measures that give equal weight to low-abundant OTUs derived from barcode 'cross-talk'.

We compared these results to those generated by a pipeline consisting of denoising followed by clustering of sequences into amplicon sequence variants with UNOISE2[52] (i.e., zero radius OTUs; zOTUs) and DADA2[51]. Quality control and trimming in UNOISE2 followed methods for UPARSE (i.e., maximum error rate of one, truncation at 170 bp; see Methods). For analyses with DADA2, we discarded all reads containing Ns or corresponding to PhiX and the remaining reads were truncated at 170 bp to match analyses using UPARSE/UNOISE2. Results from denoising and clustering into sequence variants using UNOISE2 or DADA2 were similar to results when clustering into 95% OTUs in that we observed a corresponding zOTU/ASV for each known taxon in the mock, with the exception of *Microdiplodia* sp. AK1800 and *H. polyrhiza* JEL142. The latter taxon also was missing when data were clustered with UPARSE (see Supplementary Table 9). However, UNOISE2 and DADA2 each resulted in more than one zOTU/ASV with high sequence identity to each taxon in the mock community. Because these were represented by fewer reads they likely represent spurious zOTUs/ASVs resulting from sequencing errors (see [33]). Thus, subsequent analyses were performed with OTUs resulting from UPARSE due to the lower rate of spurious errors.

We also used the mock community to assess the potential for primer bias to impact the results of our study. The primers used here successfully amplified isolates from all four phyla in the mock community (as confirmed with agarose gel electrophoresis of PCR1 products) with consistent read counts among the five replicates of the mock community. Illumina reads were recovered from 31 of 32 taxa in the normalized mock community after stringent quality control (Supplementary Table 9). Although one member of the Chytridiomycetes was amplified, we observed no reads for the other Chytridiomycetes species (*Homoloaphlyctis polyrhiza* JEL142) potentially due to mismatches in primer binding sites, differences in nuclear ribosomal DNA copy number, or the presence of a spliceosomal intron adjacent to the forward primer (see [70,71]).

We next assessed whether different fungal phyla or classes in the mock were represented by different numbers of reads, suggestive of primer bias. We observed no evidence of primer bias for Ascomycota taxa in the mock community. All taxa were represented by similar numbers of reads (ANOVA $F_{2,12} = 0.21$, $P = 0.8141$; Supplementary Table 9). In contrast, Ascomycota taxa in the mock community were represented by significantly greater read counts compared to Basidiomycota members of the mock community ($t_{27} = -2.42$, $P = 0.0224$), consistent with previous studies illustrating decreased read counts in lineages with mismatches in primer regions[70,71]. The impact of this finding will be addressed for endophyte community studies in future work.

Filtering of host reads

We used taxonomic information from MEGAN[47] LCA to filter reads representing OTUs from mycobionts of lichens (1,219,296) and plants (270,563). After filtering OTUs by taxonomy the combined culture-based and culture-free NGS data set included 6,054 OTUs (after removing 136 singleton and doubleton OTUs represented by NGS data only; see [32]). Of these OTUs, 5,668 were obtained only with NGS (4,106,890 reads), 338 were obtained from both culturing and NGS (1,828,969 reads), and 48 were found only by culturing (126 sequences).

Statistical analyses

Comparison of endophyte taxonomy for NGS and culturing

We compared the taxonomic composition of endophyte OTUs from NGS and culturing at higher taxonomic levels (phylum, class, and order) to assess the degree to which these approaches recovered the same clades. For Ascomycota, culturing and NGS recovered the same classes of Pezizomycotina, although culturing yielded only one isolate of Saccharomycotina (Supplementary Fig. 2). At the ordinal level, the combined dataset was represented by 44 orders of Ascomycota, 15 of which were found with a relative abundance > 1%. Our culturing approach detected 14 of 15 abundant orders of Ascomycota observed via NGS, failing only to isolate Chaetosphaeriales (Sordariomycetes). This contrasts with our previous isolation of numerous endophytic Chaetosphaeriales in temperate and sub-tropical plants and lichens[18].

In general, the ranked abundance of orders of Ascomycota was similar between the

culturing and NGS data sets. However, two orders appeared over-represented in the culturing dataset relative to the NGS: Xylariales and Coniochaetales. This may be due to the enhanced ability of Xylariales and Coniochaetales fungi to grow under the conditions used here, or alternatively, for primer bias to lead to poor amplification in the NGS approach (see NC0098 in Supplementary Table 9). NGS detected seven classes of Basidiomycota, three of which also were observed by culturing (Agaricomycetes, Tremellomycetes, and Exobasidiomycetes; Supplementary Fig. 2). NGS detected Chytridiomycota, Zoopagomycota, and Mucoromycota, but of these, culturing detected only the Mucoromycota (Mortierellomycotina and Mucoromycotina) (Supplementary Fig. 2).

Community structure

NMDS ordination results based on Hellinger distance were compared to those generated with an abundance-based similarity metric (Bray-Curtis) and a presence-absence distance metric (standard Jaccard). When using Bray-Curtis and Jaccard distance metrics, we transformed OTU abundances with the "metaMDS" function in the R package vegan[57] (i.e., square-root and Wisconsin double-standardization of OTU abundances). All distance measures provided very similar NMDS plots. We compared PERMANOVA results using Hellinger distance to those generated with the Bray-Curtis and Jaccard distance metrics (after transformation with "metaMDS" as described above). These distance measures provided results similar to those based on Hellinger distance.

Spatial autocorrelation and distance decay

For all analyses, variation in endophyte community composition attributable to spatial eigenvectors was negligible ($R^2 < 0.03$) and spatial eigenvectors were not a significant explanatory variable in models containing host and site. Both Hellinger distance and Bray-Curtis community dissimilarity provided similar and significant results regarding the importance of host lineage for structuring endophyte communities regardless of spatial scale.

Hierarchical clustering of endophyte communities in focal host genera

Analyses were restricted to 10 plant genera and five lichen genera sampled in a minimum of four

sites (Supplementary Fig. 7). OTUs with < 100 reads were excluded. We focused at the genus level (see Supplementary Tables 2 and 3) and read counts were rarefied to the lowest sample size per site to remove the effect of differential sampling depth (average: 16,532 reads per site per genus).

Relationship of host genetic distance and endophyte community dissimilarity

Sequences were downloaded from NCBI for each host species. When host species was unknown or sequence data for that species were not available, a representative sequence from the host genus was used as a proxy.

Species area relationships

Species area relationships (SAR) for endophytes were computed for Sanger sequences from cultures and NGS data at scales ranging from the tissue area scale (2 mm^2 per tissue piece sampled) to the sampling area (ca. 75 m^2 per microsite), the continental scale, and the final scope of the study (Supplementary Fig. 8, Supplementary Table 7). To calculate the number of observed species in > 1 site (i.e., area > 225 m^2, the sum of three microsites per site), we calculated the mean richness of all possible pairwise combinations of two sites, three sites, four sites, etc. For estimates based on the area of photosynthetic tissues, we calculated the mean number of species within a host individual (i.e., 96 tissue pieces), within a host species (96 tissue pieces x 3 microsites), within a microsite (96 tissue pieces x number of host individuals in a microsite), and within a site (96 tissue pieces x number of host individuals in a site). The relationship between species and area was highly significant with both culture-free and NGS data, and at different scales (Supplementary Table 7). From the regression we examined the slope of the line, which describes the rate of increase in richness with area (z-value). Similar z-values were obtained when SAR were evaluated at different scales and using different endophyte sampling methods (i.e., culturing and culture-free NGS; see Supplementary Table 7).

Endophyte host associations

We quantified and visualized the distribution of OTUs among major host lineages with networks that were constructed via a script developed in R with the package igraph 0.7.1[68]. Networks were

constructed for OTUs in each site (using endophytes from all host taxa; Supplementary Tables 2-3). For each site, data for all species within a host lineage were combined (see Supplementary Tables 2-3) and read counts were rarefied to the minimum number of reads per lineage in that site (e.g., 25,784 reads per host lineage in SWM). Combining and rarefying data by lineage resulted in slightly different values per site compared to NMDS analyses done for Supplementary Fig. 1, where reads were rarefied to the depth in each site as a function of host taxon. 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 (Supplementary Fig. 7) or (2) endophyte communities from a representative genus for each major host lineage (Fig. 3). Computing networks with genera found in a minimum of four sites reduced network complexity for visualization while also accounting for the significant effect of host genus on endophyte community structure (see Supplementary Fig. 1). Network structures at the circumboreal scale were similar regardless of taxon selection within a host lineage (e.g., use of *Rhododendron* resulted in similar outcomes as other Magnoliophyta). For networks computed for hosts within sites, we removed OTUs with < 100 reads across the entire dataset to reduce the complexity of networks for visualization (Fig. 3a to g). For networks computed at the circumglobal scale, we removed OTUs with < 300 reads across the entire dataset to reduce network complexity for visualization (Fig. 3h to j, Supplementary Fig. 9). Relative read abundance of common OTUs used in the networks was a poor predictor of the number of host lineages in which an OTU was found ($R^2 = 0.09$; also see Supplementary Fig. 9): log relative read abundance for OTUs that occurred one, two, three, or four host lineages did not differ significantly (ANOVA; $F_{3, 496} = 2.08$, $P = 0.1013$). However, OTUs found in ≥ 5 host lineages had significantly more reads than OTUs found in < 5 host lineages (t-test, $t_{659} = 8.25$, $P < 0.001$). Thus, endophytes found in a small number of hosts were not represented by fewer reads such that specificity was not confounded by rarity.

Supplementary References (listed only in Supplementary Methods and Supplementary Table 3)

70. Tedersoo, L. & Lindahl, B. Fungal identification biases in microbiome projects. *Environ. Microbiol. Rep.* (2016). doi:10.1111/1758-2229.12438

71. Bellemain, E. *et al.* ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol.* **10**, 189 (2010).

72. U'Ren, J. M. *et al.* Diversity and evolutionary origins of fungi associated with seeds of a neotropical pioneer tree: a case study for analysing fungal environmental samples. *Mycol. Res.* **113**, 432–449 (2009).

73. Lücking, R., Hodkinson, B. P. & Leavitt, S. D. The 2016 classification of lichenized fungi in the Ascomycota and Basidiomycota – Approaching one thousand genera. *Bryologist* **119**, 361–416 (2016).

74. Brodo, I. M., Sharnoff, S. D., Sharnoff, S. & Canadian Museum of Nature. *Lichens of North America*. (Yale University Press, 2001).

75. Piercey-Normore, M. D. & Depriest, P. T. Algal switching among lichen symbioses. *Am. J. Bot.* **88**, 1490–1498 (2001).

76. Blaha, J., Baloch, E. & Grube, M. High photobiont diversity associated with the euryoecious lichen-forming ascomycete *Lecanora rupicola* (Lecanoraceae, Ascomycota). *Biol. J. Linn. Soc. Lond.* **88**, 283–293 (2006).

77. Hauck, M., Helms, G. & Friedl, T. Photobiont selectivity in the epiphytic lichens *Hypogymnia physodes* and *Lecanora conizaeoides*. *Lichenologist* **39**, 195–204 (2007).

78. Bačkor, M., Peksa, O., Škaloud, P. & Bačkorová, M. Photobiont diversity in lichens from

metal-rich substrata based on ITS rDNA sequences. *Ecotoxicol. Environ. Saf.* **73**, 603–612 (2010).

79. Thüs, H. *et al.* Revisiting photobiont diversity in the lichen family Verrucariaceae (Ascomycota). *Eur. J. Phycol.* **46**, 399–415 (2011).
80. He, Y. & Zhang, Z. Diversity of organism in the *Usnea longissima* lichen. *Afr. J. Microbiol. Res.* **6**, 4797–4804 (2012).
81. Dal Grande, F., Beck, A., Cornejo, C. & Singh, G. Molecular phylogeny and symbiotic selectivity of the green algal genus *Dictyochloropsis* sl (Trebouxiophyceae): a polyphyletic and widespread group forming photobiont-mediated guilds in the lichen family Lobariaceae. *New Phytologist* **202**, 455–470 (2014).
82. Muggia, L., Pérez-Ortega, S., Kopun, T., Zellnig, G. & Grube, M. Photobiont selectivity leads to ecological tolerance and evolutionary divergence in a polymorphic complex of lichenized fungi. *Ann. Bot.* **114**, 463–475 (2014).
83. Řídká, T., Peksa, O., Rai, H., Upreti, D. K. & Škaloud, P. Photobiont diversity in Indian *Cladonia* lichens, with special emphasis on the geographical patterns. in *Terricolous Lichens in India: Volume 1: Diversity Patterns and Distribution Ecology* (eds. Rai, H. & Upreti, D. K.) 53–71 (Springer New York, 2014).
84. Lindgren, H. *et al.* High fungal selectivity for algal symbionts in the genus *Bryoria*. *Lichenologist* **46**, 681–695 (2014).
85. Cao, S. *et al.* Distribution patterns of haplotypes for symbionts from *Umbilicaria esculenta* and *U. muehlenbergii* reflect the importance of reproductive strategy in shaping population genetic structure. *BMC Microbiol.* **15**, 212 (2015).
86. Cornejo, C. & Scheidegger, C. Multi-gene phylogeny of the genus *Lobaria*: Evidence of

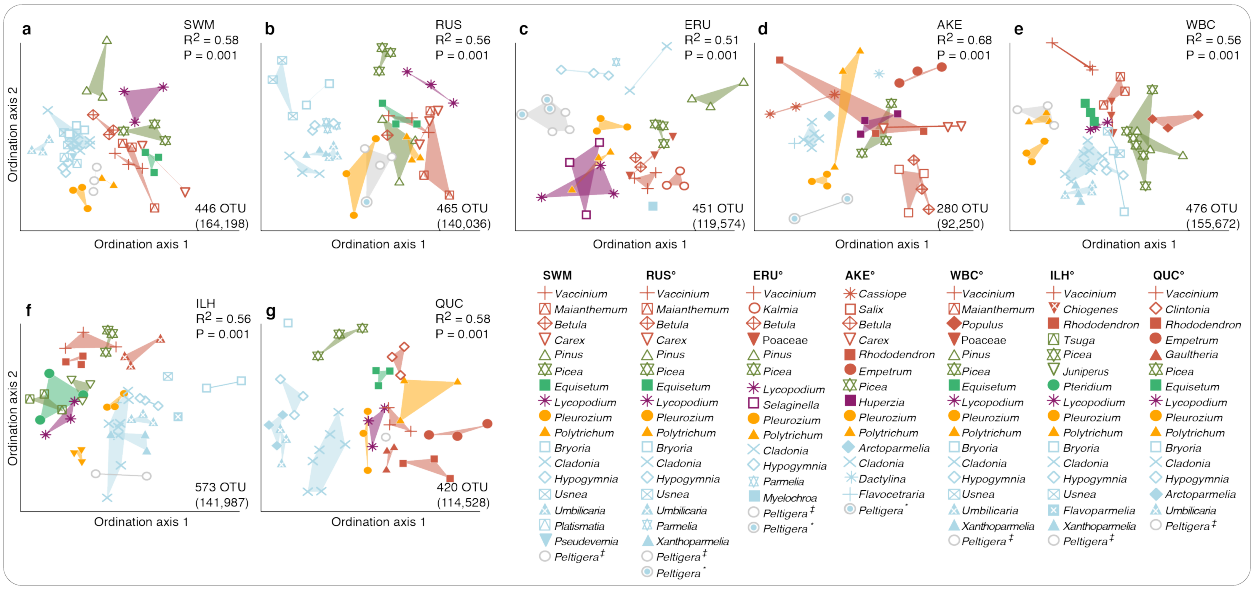
species-pair and allopatric cryptic speciation in East Asia. *Am. J. Bot.* **102**, 2058–2073 (2015).

87. Magain, N., Miadlikowska, J., Goffinet, B., Sérusiaux, E. & Lutzoni, F. Macroevolution of specificity in cyanolichens of the genus *Peltigera* section Polydactylon (Lecanoromycetes, Ascomycota). *Syst. Biol.* (2016).

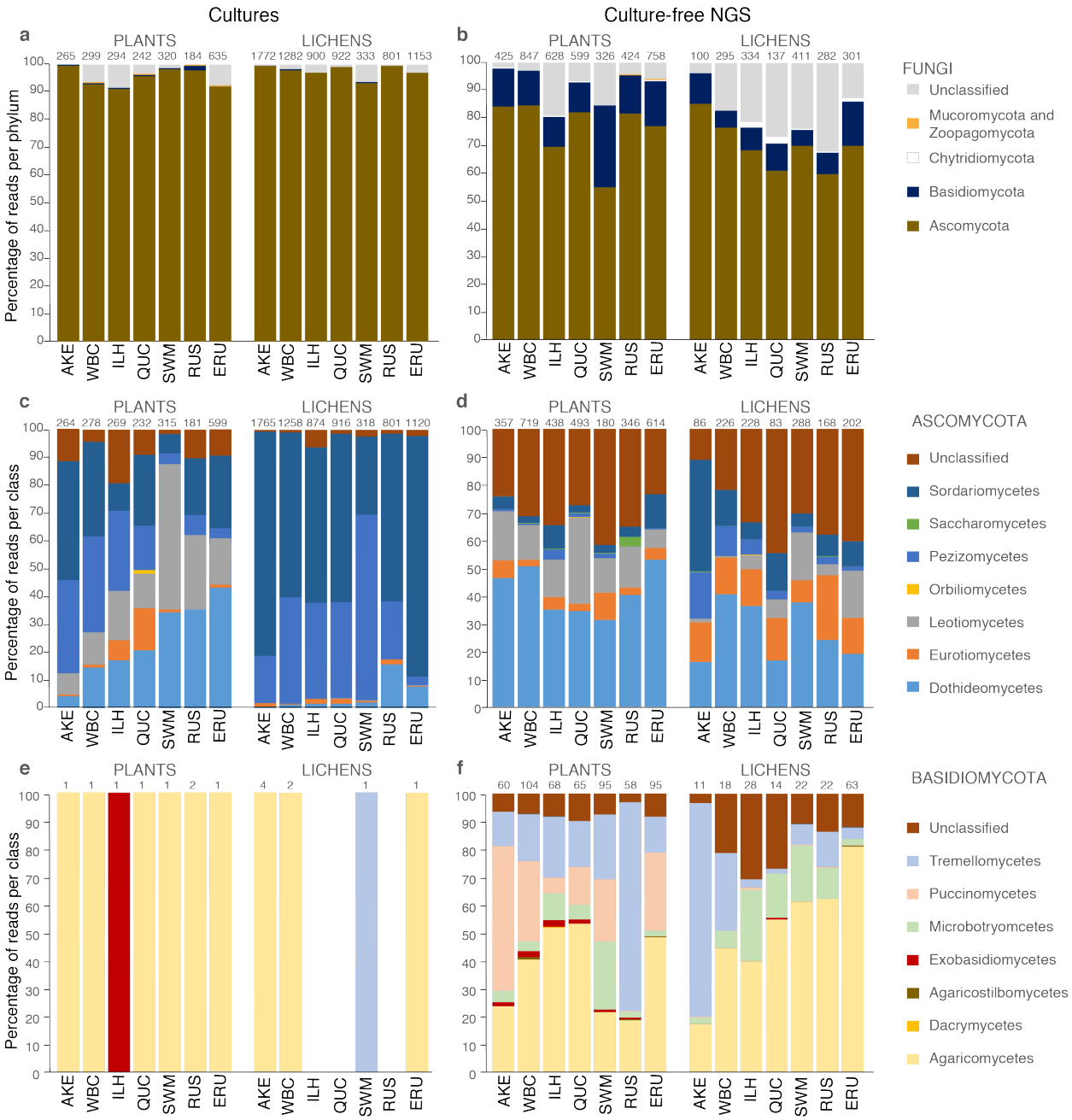
88. Magain, N. *et al.* Conserved genomic collinearity as a source of broadly applicable, fast evolving, markers to resolve species complexes: A case study using the lichen-forming genus *Peltigera* section Polydactylon. *Mol. Phylogenet. Evol.* **117**, 10–29 (2017).

89. Škaloud, P., Friedl, T., Hallmann, C., Beck, A. & Dal Grande, F. Taxonomic revision and species delimitation of coccoid green algae currently assigned to the genus *Dictyochloropsis* (Trebouxiophyceae, Chlorophyta). *J. Phycol.* **52**, 599–617 (2016).

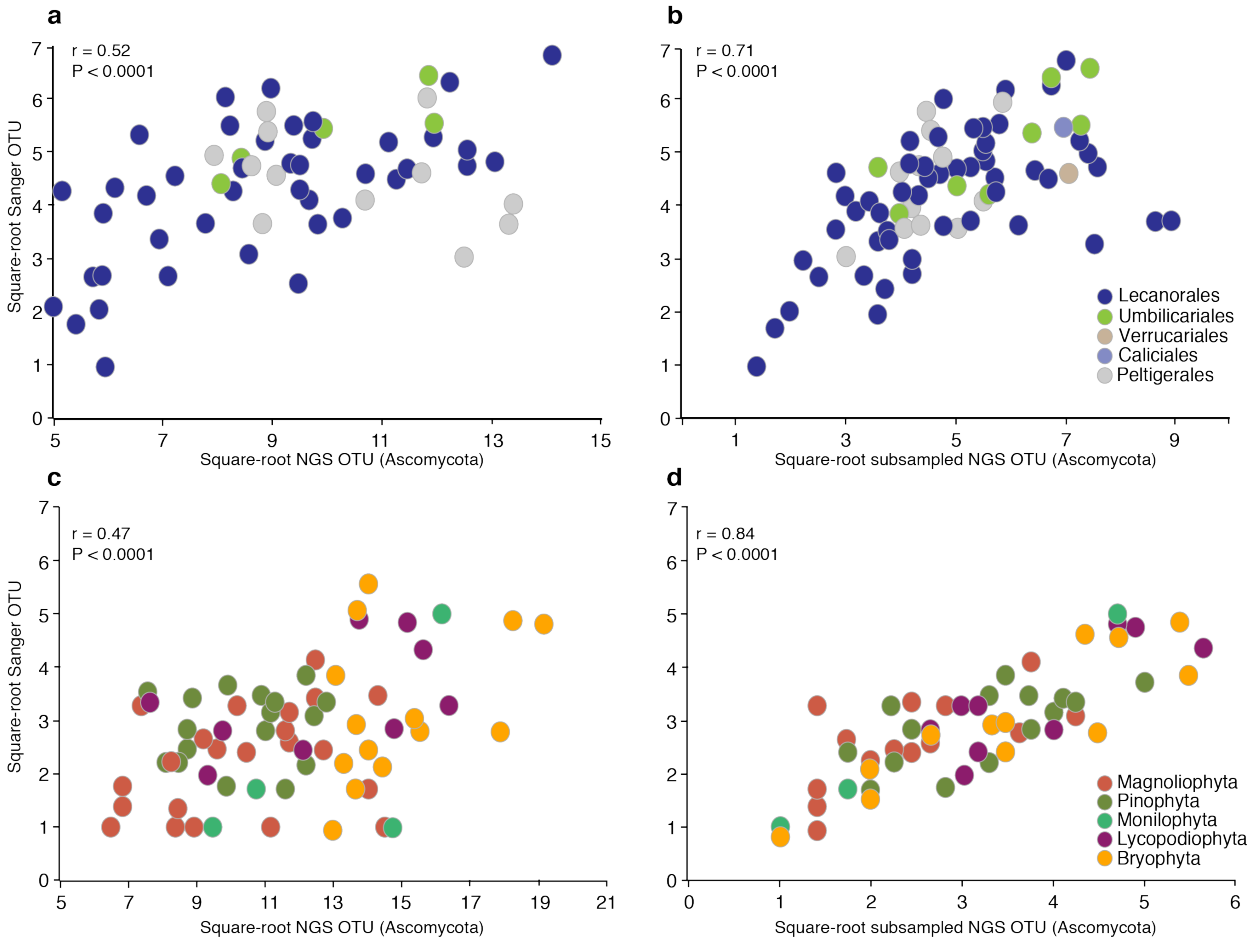
Supplementary Figures



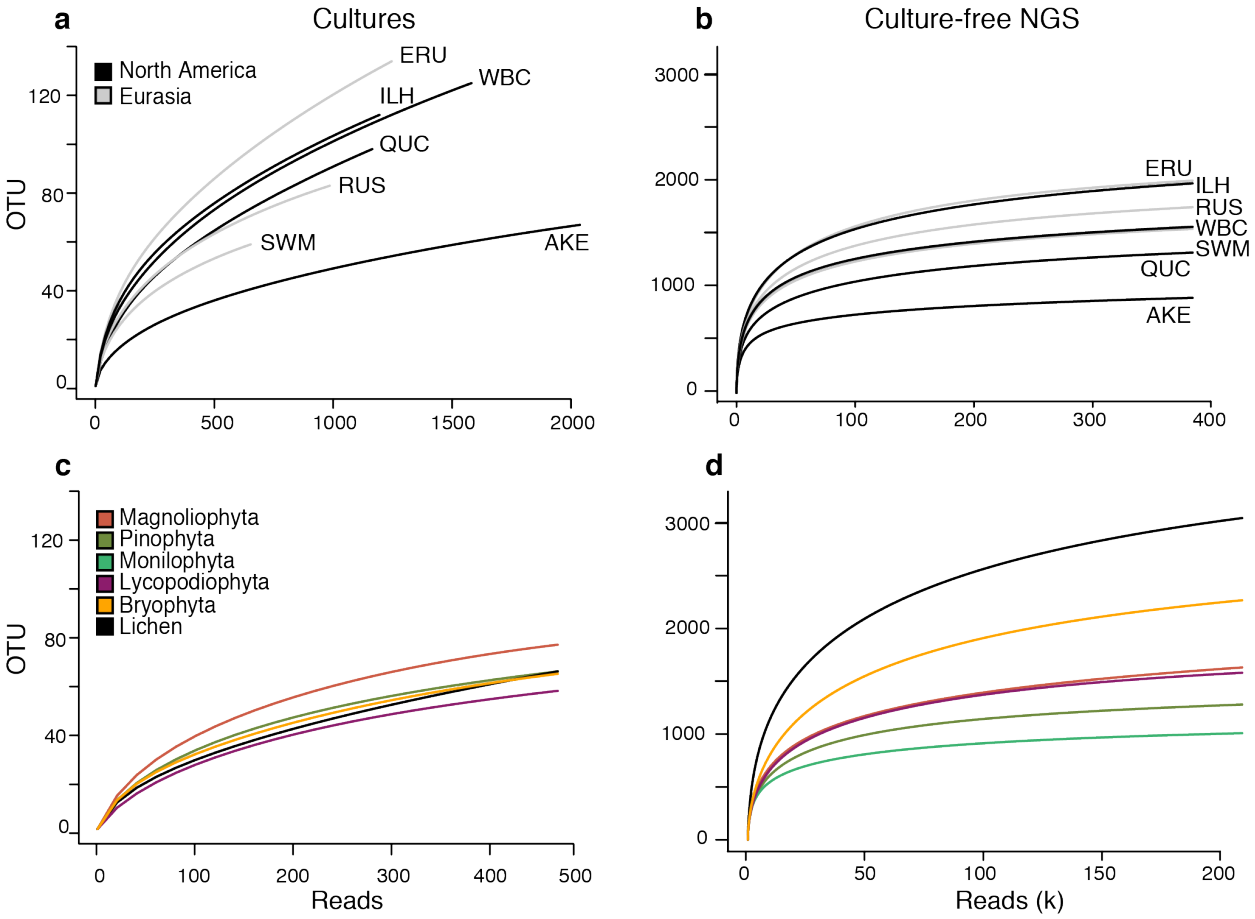
Supplementary Fig. 1. Host identity was the major predictor of endophyte community structure in seven boreal forests. a-g, Non-metric multidimensional scaling (NMDS) of endophyte assemblages in each site (Supplementary Table 1). Richness and read number are in the lower right of each. Color indicates host lineage, as in Fig. 1; shape and convex hulls indicate host genera (lower right). Statistics are from permutational multivariate analyses of variance (PERMANOVA). Reads were rarefied to the same sampling depth per host. Hosts with < 2,250 reads and OTUs with < 100 reads were excluded. ° indicates that a subset of hosts were excluded from analyses due to insufficient read depth (see Supplementary Tables 2 and 3). For *Peltigera*, photobionts include Cyanobacteria, which can occur alone within thalli (‡) or together with a green alga within each thallus, i.e., tripartite lichens.



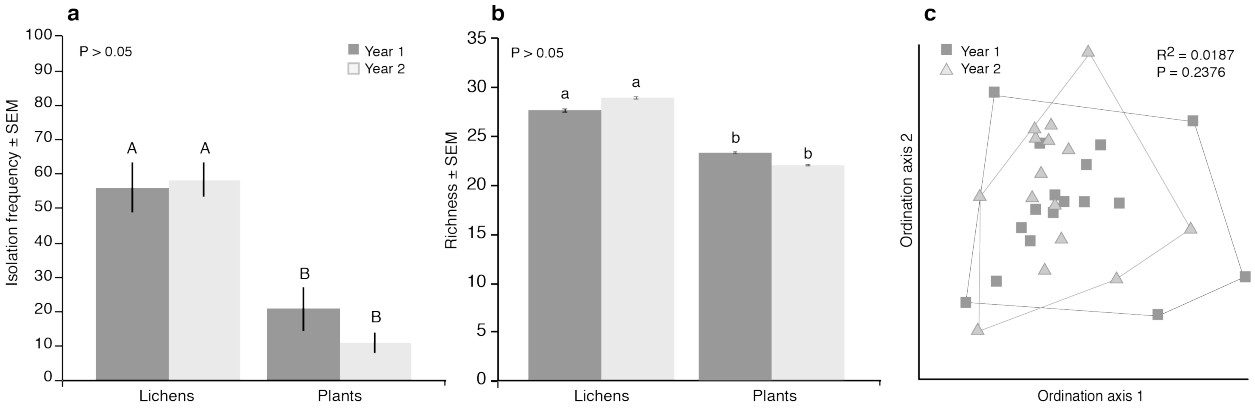
Supplementary Fig. 2. Taxonomy of fungal endophytes in seven sites across the circumboreal belt as a function of site and host. Fungal taxonomy is shown at the phylum level for **a**, Sanger sequencing of cultures and **b**, culture-free NGS. Class-level taxonomy (based on ITS2 region) is shown for Ascomycota as observed by **c**, cultures and **d**, NGS, and for Basidiomycota as observed by **e**, cultures and **f**, NGS. Site names are abbreviated following Supplementary Table 1. Sample size is shown at the top of each bar (culture-free NGS reads are shown in thousands). Sanger sequences without the ITS2 region were excluded from this analysis (see Supplementary Tables 2 and 3).



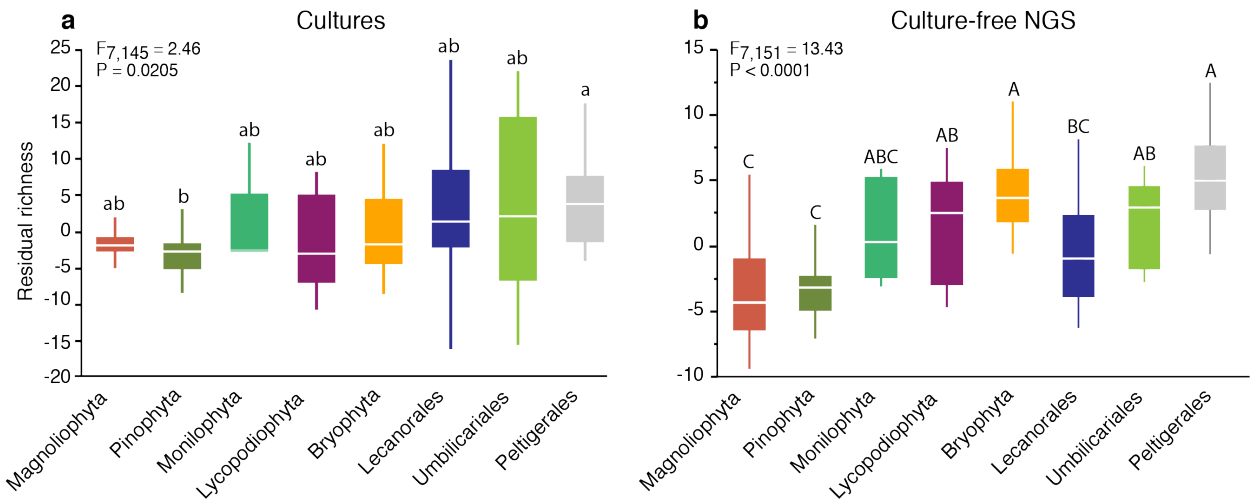
Supplementary Fig. 3. Significant correlation between estimates of endophyte richness based on Sanger sequences of cultures and culture-free NGS for seven boreal sites. Points are colored according to host lineage for lichens (**a** and **b**) and plants (**c** and **d**). To account for differences in sequencing depth, NGS reads for each host lineage/site were subsampled to a minimum read depth (**a**, 3,106 per host individual for lichens; **c**, 21,203 per host individual for plants; Supplementary Tables 2 and 3) or subsampled to match the depth of Sanger sequencing for each host individual (**b, d**). OTU counts were square-root transformed prior to statistical analysis. Outliers, host individuals with zero Sanger reads (n = 4), and/or read levels below the minimum threshold were removed from all analyses.



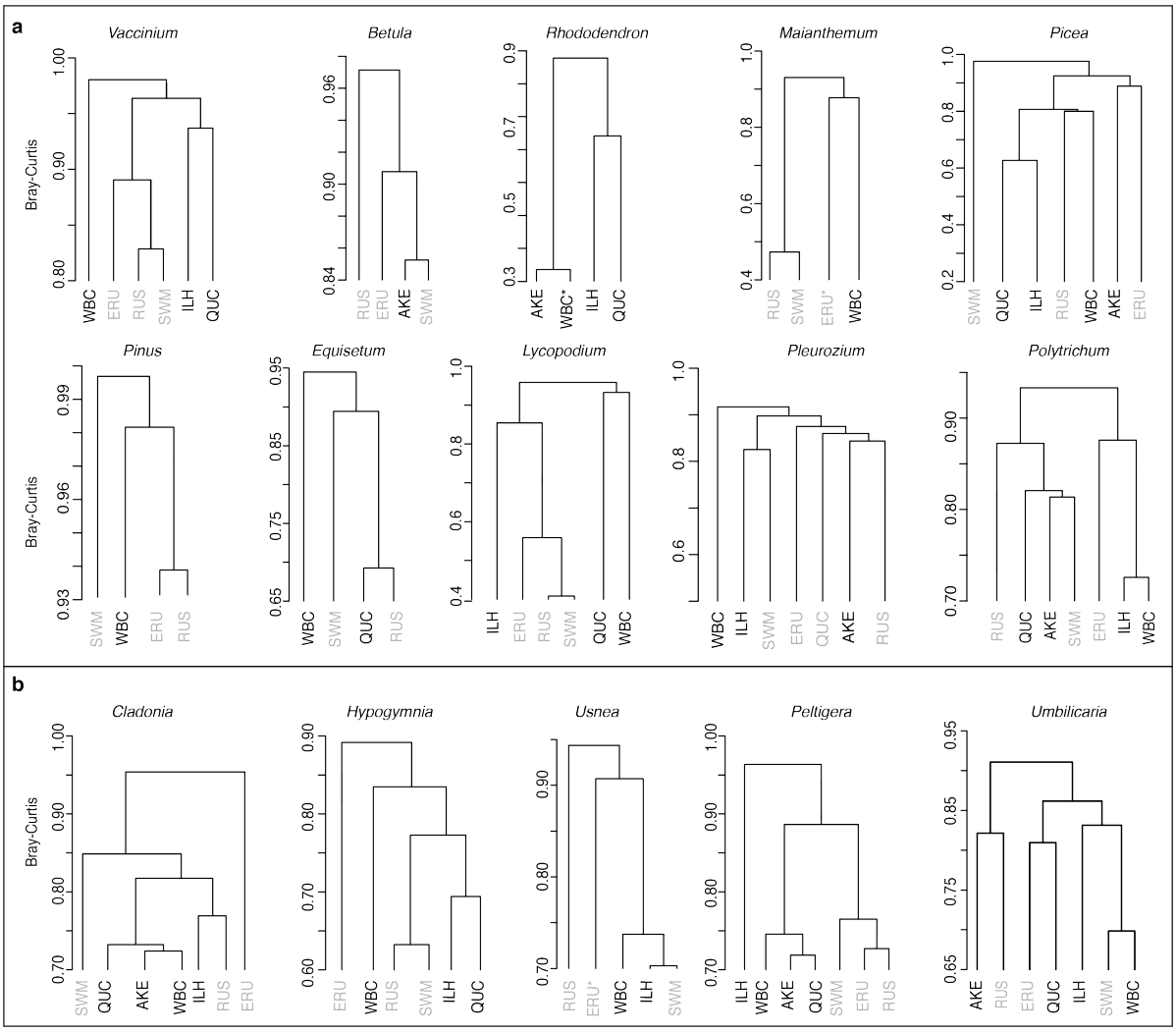
Supplementary Fig. 4. Rarefaction curves for fungal endophytes as a function of site and host. Each line represents the individual-based rarefaction curve for endophytic OTUs from each site (**a** and **b**) or host lineage (**c** and **d**) based on culturing and culture-free NGS. Site names are abbreviated following Supplementary Table 1. Rarefaction sequencing depth is based on the minimum number of sequences per site or host lineage. Monilophyta were excluded from panel c due to low numbers of cultured endophytes.



Supplementary Fig. 5. Fungal endophyte communities from two sampling events in a focal site (Eagle Summit, Alaska: AKE). Analyses indicate consistent **a,** isolation frequency; **b,** richness; and **c,** community composition. Data from the first year are based on results in [18]. In **a** and **b**, darker colors indicate sampling in the first year (2008); lighter colors indicate sampling in the second year (2011). Bars with the same letter within a panel indicate no statistical differences following a t-test of isolation frequency (uppercase letters) or richness (lowercase letters). In **b**, data were rarefied 1,000 times to the same depth and the mean and standard error were calculated. **c,** NMDS analysis of endophyte communities from the same hosts in two sampling years (2008, 2011). Statistics represent results of PERMANOVA examining the effect of sampling year.

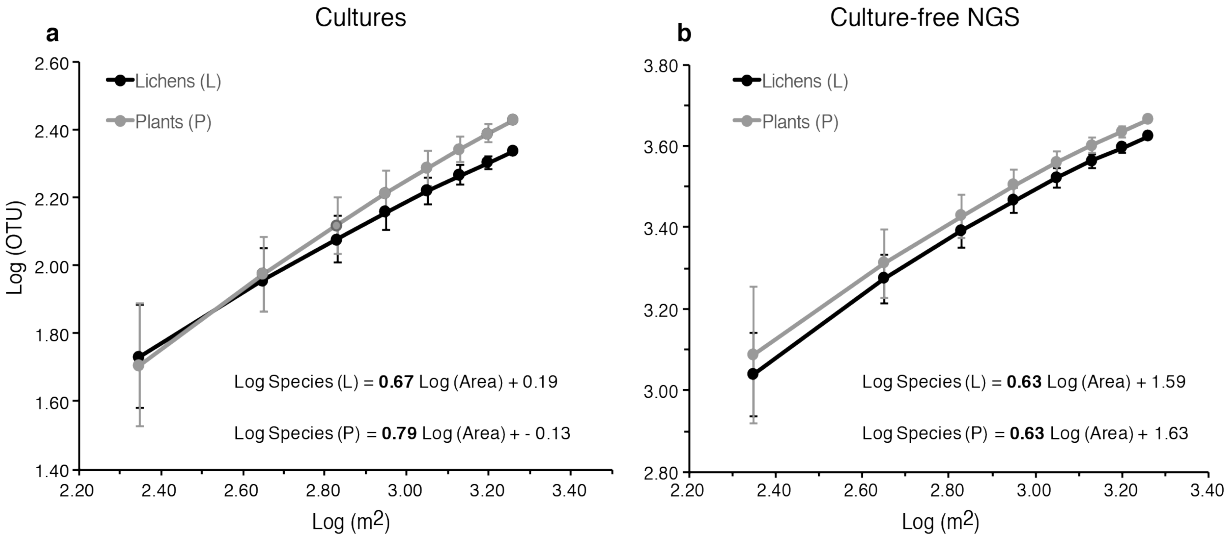


Supplementary Fig. 6. Fungal endophyte richness varied among host lineages, as inferred by **a**, cultures and **b**, culture-free NGS. One-way ANOVA was used to compare residuals of OTU richness for cultures and NGS in relation to the square-root of the number of reads to account for differences in sequencing depth[22]. Letters indicate statistical differences in richness based on post-hoc analysis (Tukey's HSD) (cultures: lowercase; culture-free NGS: uppercase). Data from all hosts (including both main and secondary sites, Supplementary Tables 2 and 3) were used for comparisons of richness with the exception of outliers and mycobiont orders with a single individual (Supplementary Table 3). Quantile box plots show the minimum and maximum values (vertical lines), the 50th percentile (i.e., median; white line), and the 25th to 75th percentiles (area of the box).

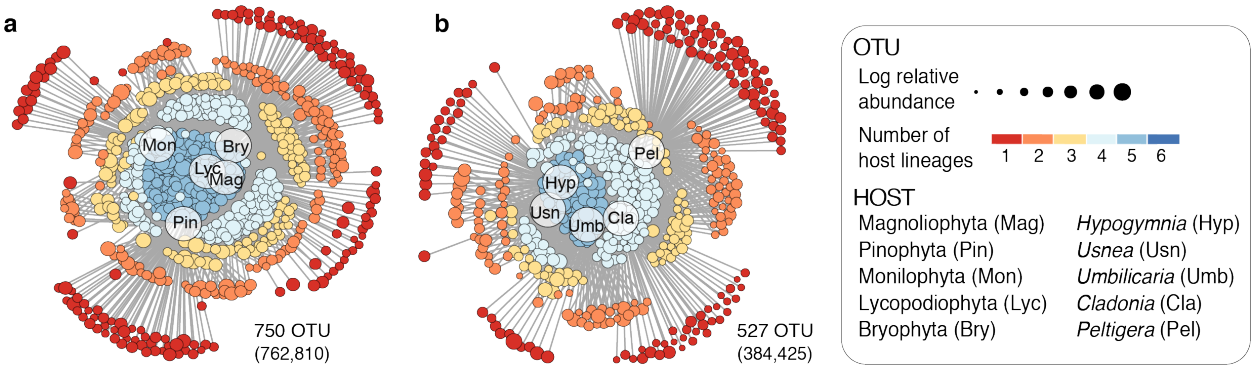


Supplementary Fig. 7. Site-level clustering of fungal endophyte communities for focal host genera illustrates distinct patterns for each genus that do not reflect geographic distance between sites. Hierarchical clustering is based on UPGMA average linkage with the Bray-Curtis metric for **a,** plants and **b,** lichens. OTUs with < 100 reads and host species sampled in fewer than four sites were excluded. We analyzed the data at the genus level for each site (Supplementary Tables 2 and 3) and read counts were rarefied to the lowest sample size per site. We used an average of 16,532 reads per site per genus. In some cases, combining data for multiple species within a site resulted in the inclusion of genera that were excluded previously from NMDS analyses (Supplementary Fig. 1) due to too few reads. Site names are abbreviated following Supplementary Table 1 and shaded according to continent: North America (black) and Eurasia (grey). Asterisks (*) indicate host taxa that were collected from secondary sampling sites

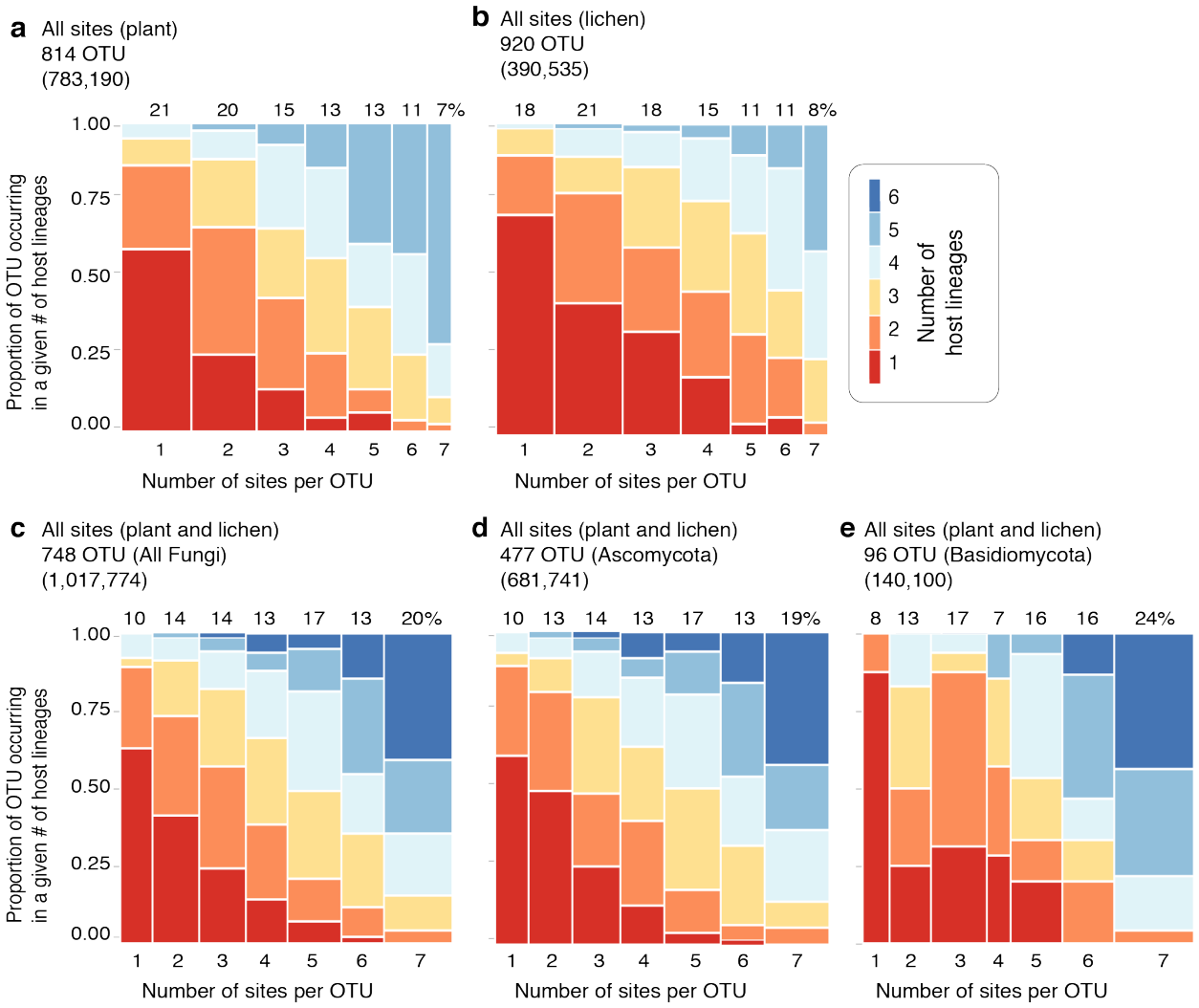
(Supplementary Tables 2 and 3), which were used here to provide the greatest number of genus-level comparisons across sites.



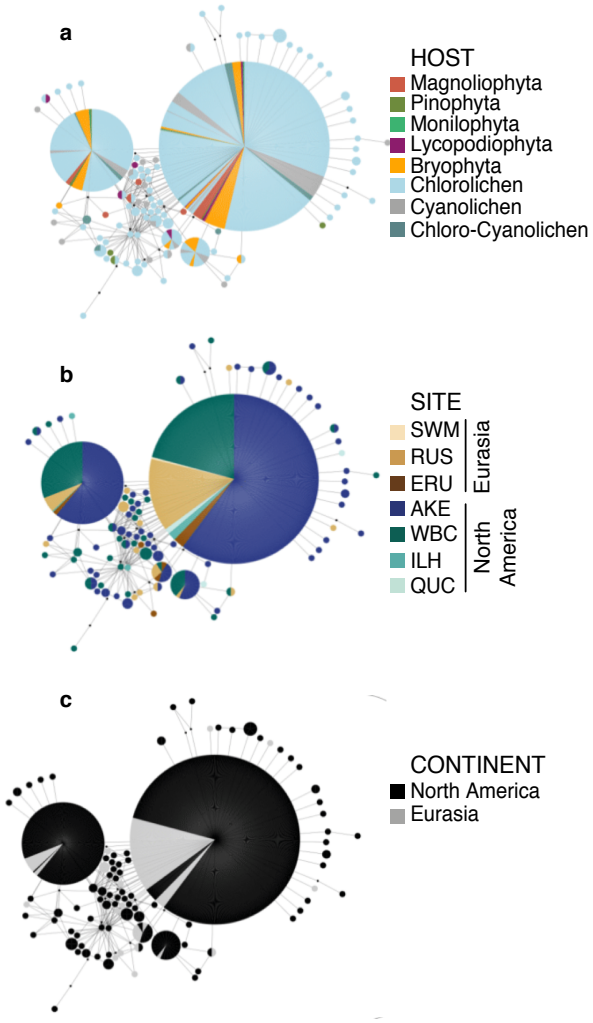
Supplementary Fig. 8. Species area relationships (SAR) for fungal endophytes are steep and similar for plant and lichen hosts, as determined from **a**, cultures and **b**, NGS (cultures, lichens: $R^2 = 0.71$, $P < 0.0001$; cultures, plants: $R^2 = 0.84$, $P < 0.0001$; NGS, lichens: $R^2 = 0.82$, $P < 0.0001$; NGS, plants: $R^2 = 0.89$, $P < 0.0001$). The slope of the line (z-value) describes the rate of increase in richness with area. Similar z-values are derived when SAR are evaluated within continents only (see Supplementary Table 7). Solid circles and vertical bars represent the mean and standard deviation of the mean, respectively.



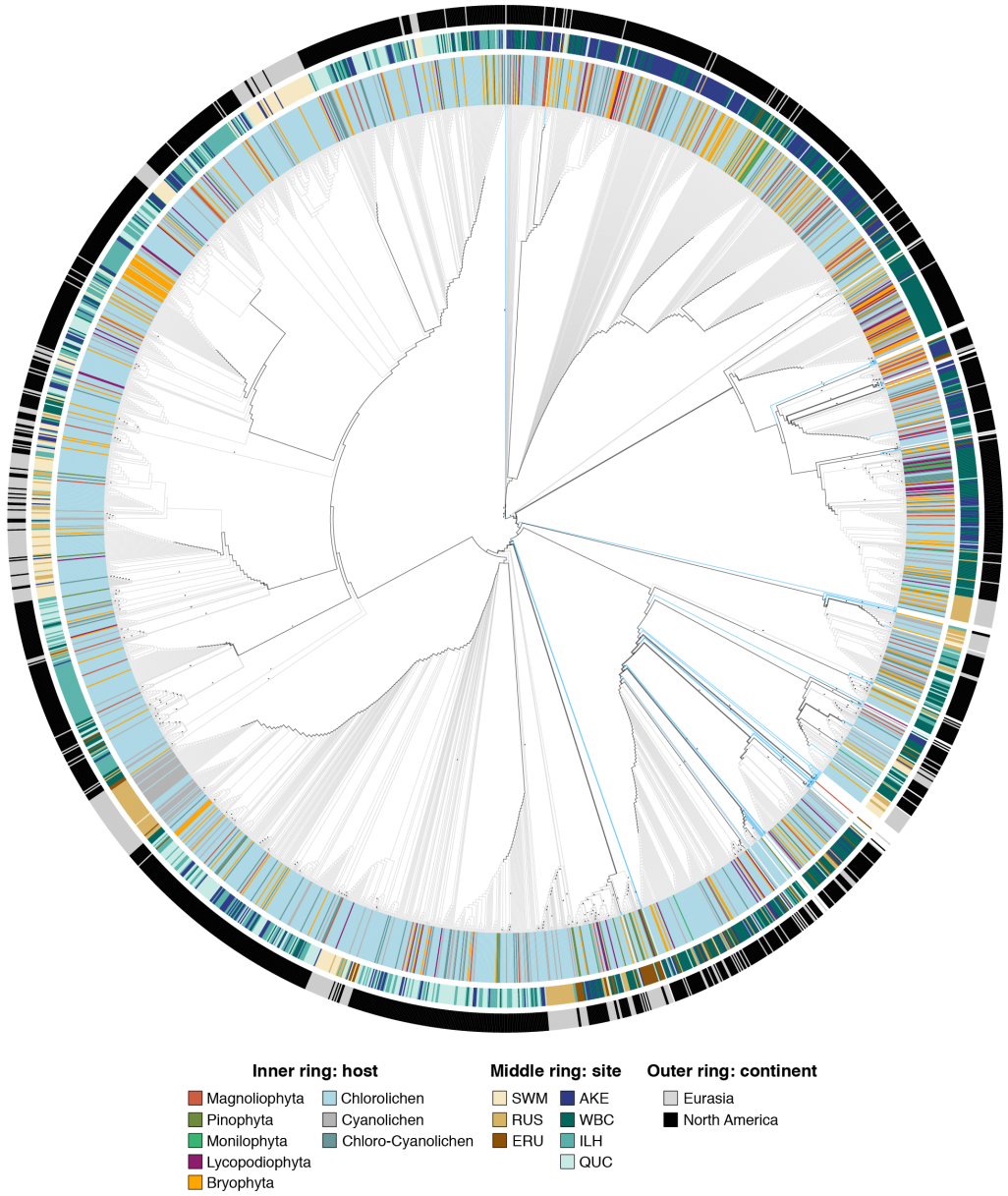
Supplementary Fig. 9. Global networks of fungal endophyte communities as a function of host lineage in seven sites across the circumboreal belt reveals no correlation between read abundance and occurrence in different host lineages. Each node represents an OTU. Edges connect OTUs to host lineage(s) in which they were found. Color indicates the number of host lineages and node diameter is proportional to OTU $\log_{10}$ relative read abundance. OTU richness and total read counts are shown above each network. **a**, Network based on culture-free NGS data for plant phyla based on ten host genera that were sampled in a minimum of four sites: *Rhododendron, Vaccinium, Betula,* and *Maianthemum* (Magnoliophyta, Mag), *Picea* and *Pinus* (Pinophyta, Pin), *Equisetum* (Monilophyta, Mon), *Lycopodium* (Lycopodiophyta, Lyc), and *Pleurozium* and *Polytrichum* (Bryophyta, Bry). OTUs with < 300 reads were excluded to improve network visualization. **b**, Network based on culture-free NGS data for lichens. Circles and read number are as above. Lichen mycobionts are represented by five genera that were sampled in a minimum of four sites: *Peltigera* (Pel), *Umbilicaria* (Umb), *Hypogymnia* (Hyp), *Usnea* (Usn), and *Cladonia* (Cla). OTUs with < 300 reads were excluded to improve network visualization. Each lichen mycobiont forms symbioses with Chlorophyta photobiont(s), with the exception of some *Peltigera* mycobionts that associate solely with cyanobacterial photobionts or with both photobionts simultaneously (Supplementary Table 3).



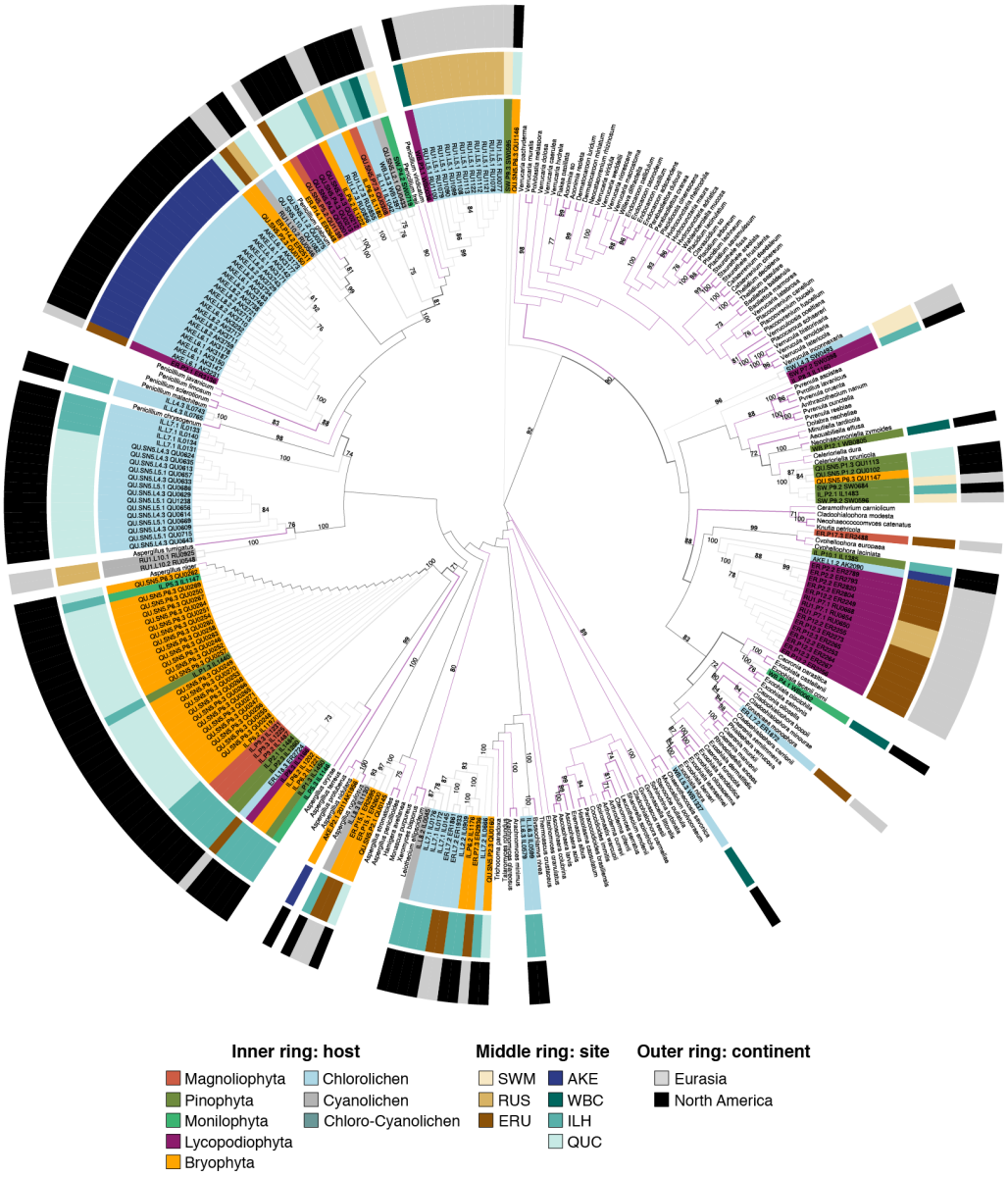
Supplementary Fig. 10. The relationship of host breadth and geographic breadth for boreal endophytes. The proportion of OTUs occurring in a given number of distinct host lineages (i.e., one, two, three, four, five, or six, denoted by color) as a function of the number of sites in which that OTU was found. **a**, All OTUs from plants (corresponding to network Fig. 3h), **b**, all OTUs from lichens (corresponding to network Fig. 3i), **c**, all OTUs from lichen and plant hosts (corresponding to network Fig. 3j), **d**, Ascomycota OTUs from lichen and plant hosts, **e**, Basidiomycota OTUs from lichen and plant hosts. The percentage of OTUs per site/host lineage is shown at the top of each bar. For each panel, we found a significant relationship between the number of host lineages and the number of sites in which an OTU was found (likelihood ratio tests; P-value < 0.001 for all networks).



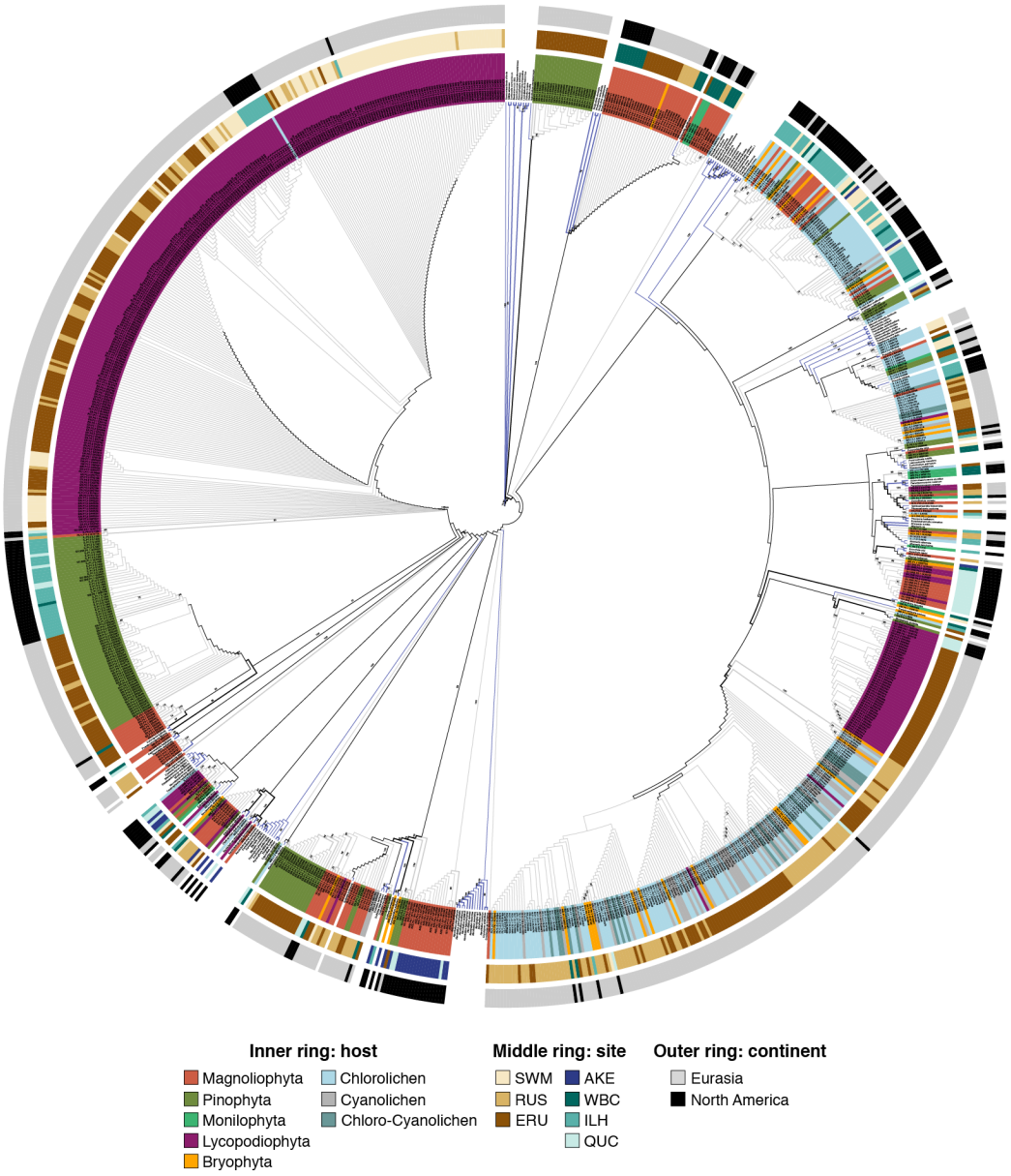
Supplementary Fig. 11. Analysis of the most widespread generalist endophyte OTU (*Daldinia loculata*) reveals global distribution of the most abundant haplotypes. The network is colored as a function of **a**, host; **b**, geographic location; and **c**, continent. Chlorolichen: lichens with Chlorophyta photobionts; Cyanolichen: lichens with cyanobacterial photobionts; Chloro-Cyanolichen: lichens with both cyanobacterial and Chlorophyta photobionts in the same thallus (i.e., tripartite lichens; Supplementary Table 3). The size of each node in the network is proportional to the frequency of the haplotype in the sample. A line separating any two nodes represents a point mutation. Site names are abbreviated following Supplementary Table 1.



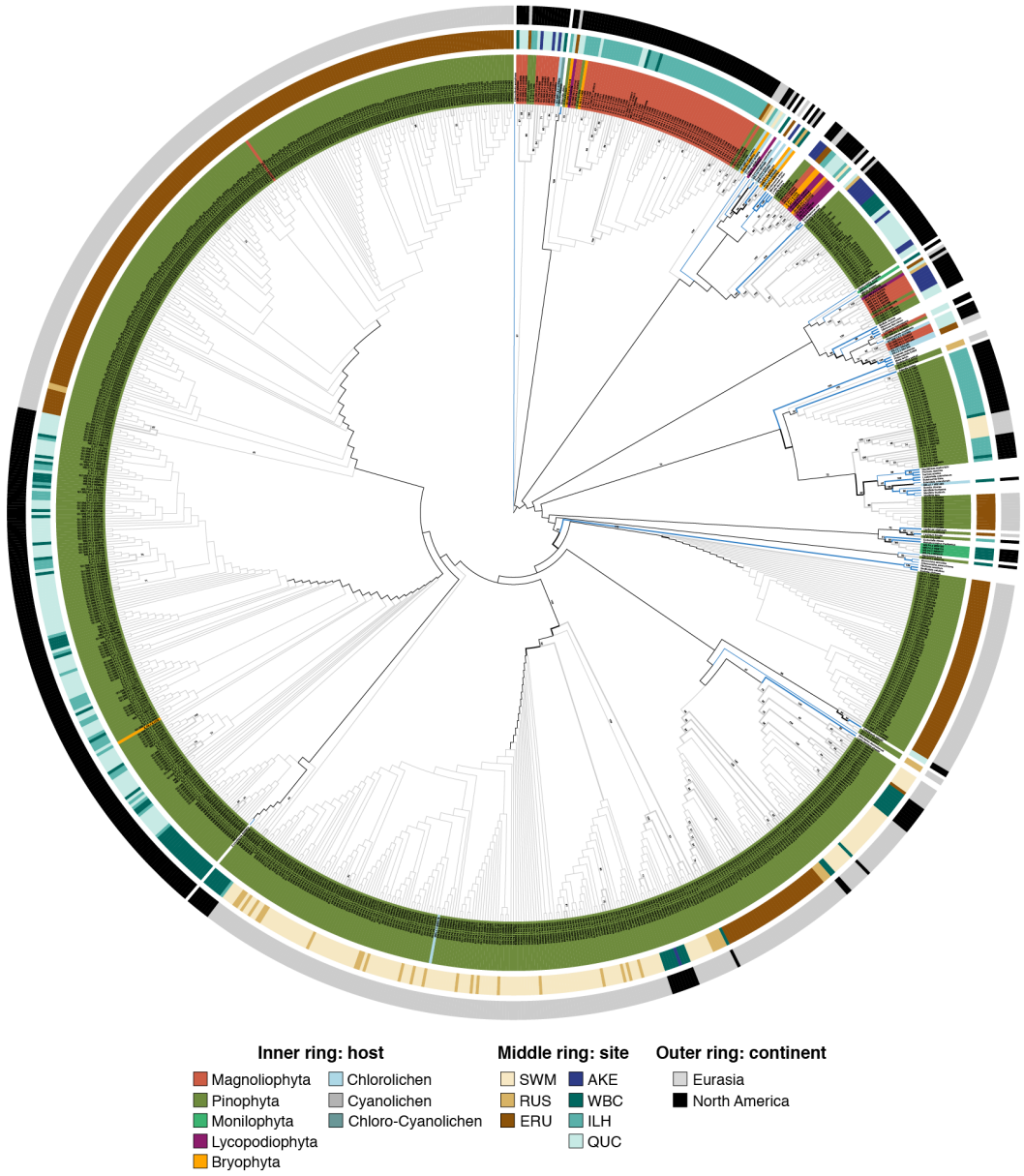
Supplementary Fig. 12. Phylogenetic placement of cultured endophytes in the evolutionary context of Pezizomycetes (Ascomycota) based on 5.8S nrDNA and LSU nrDNA as implemented in T-BAS[30]. Reference taxa are shown with colored branches and no metadata[30]. Endophytes are shown with black branches and colored rings of metadata (host, site, continent). Support values represent 1000 bootstrap replicates. Legend follows Fig. 4.



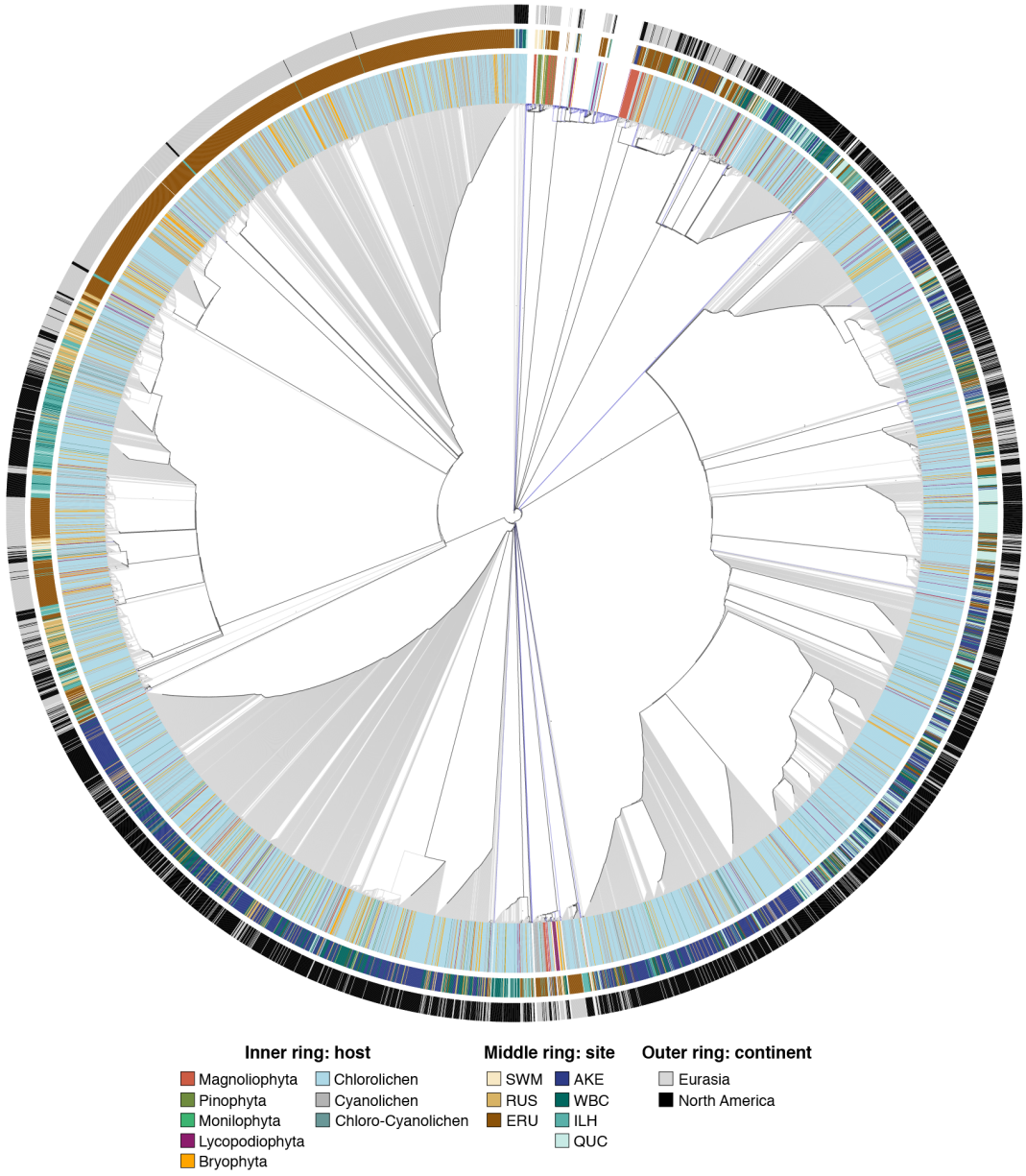
Supplementary Fig. 13. Phylogenetic placement of cultured endophytes in the evolutionary context of Eurotiomycetes (Ascomycota) based on 5.8S nrDNA and LSU nrDNA as implemented in T-BAS[30]. Reference taxa are shown with colored branches and no metadata[30]. Endophytes are shown with black branches and colored rings of metadata (host, site, continent). Support values represent 1000 bootstrap replicates. Legend follows Fig. 4.



Supplementary Fig. 14. Phylogenetic placement of cultured endophytes in the evolutionary context of Dothideomycetes (Ascomycota) based on 5.8S nrDNA and LSU nrDNA as implemented in T-BAS[30]. Reference taxa are shown with colored branches and no metadata[30]. Endophytes are shown with black branches and colored rings of metadata (host, site, continent). Support values represent 1000 bootstrap replicates. Legend follows Fig. 4.



Supplementary Fig. 15. Phylogenetic placement of cultured endophytes in the evolutionary context of Leotiomycetes (Ascomycota) based on 5.8S nrDNA and LSU nrDNA as implemented in T-BAS[30]. Reference taxa are shown with colored branches and no metadata[30]. Endophytes are shown with black branches and colored rings of metadata (host, site, continent). Support values represent 1000 bootstrap replicates. Legend follows Fig. 4.



Supplementary Fig. 16. Phylogenetic placement of cultured endophytes in the evolutionary context of Sordariomycetes (Ascomycota) based on 5.8S nrDNA and LSU nrDNA as implemented in T-BAS[30]. Reference taxa are shown with colored branches and no metadata[30]. Endophytes are shown with black branches and colored rings of metadata (host, site, continent). Support values represent 1000 bootstrap replicates. Legend follows Fig. 4.