Report

Current Biology

Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome

Highlights

- Endophyte associations with boreal plants and lichens are defined by climate
- Effects of climate on endophyte diversity differ among hosts and fungal lineages
- Endophyte biodiversity hotspots occur at key plant-community transitions
- Climate change threatens symbiotic fungi in Earth's largest terrestrial biome

Authors

Jana M. U'Ren, Shuzo Oita, François Lutzoni, ..., Denis Valle, Valerie Trouet, A. Elizabeth Arnold

Correspondence

arnold@ag.arizona.edu

In brief

U'Ren and Oita et al. use comprehensive, biome-scale sampling to show how hidden fungal symbionts of plants and lichens, and their biodiversity hotspots, are structured by climate—thus highlighting the potential sensitivity of foundational symbioses in Earth's largest terrestrial biome to the effects of climate change.

U'Ren et al., 2024, Current Biology *34*, 1–9
March 11, 2024
https://doi.org/10.1016/j.cub.2024.01.063

CellPress

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

Current Biology

CellPress

Report

Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome

Jana M. U'Ren,[1,11] Shuzo Oita,[2,11] François Lutzoni,[3] Jolanta Miadlikowska,[3] Bernard Ball,[3,4] Ignazio Carbone,[5] Georgiana May,[6] Naupaka B. Zimmerman,[7] Denis Valle,[8] Valerie Trouet,[9] and A. Elizabeth Arnold[2,10,12,*]

[1]Department of Plant Pathology, Washington State University, Pullman, WA 99164, USA
[2]School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
[3]Department of Biology, Duke University, Durham, NC 27708, USA
[4]School of Biology and Environmental Science, University College Dublin, Science Centre Belfield, Dublin D04 V1W8, Ireland
[5]Center for Integrated Fungal Research, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695, USA
[6]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108, USA
[7]Department of Biology, University of San Francisco, San Francisco, CA 94117, USA
[8]School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, FL 32611, USA
[9]Laboratory of Tree Ring Research, University of Arizona, Tucson, AZ 85721, USA
[10]Department of Ecology and Evolutionary Biology, BIO5 Institute, Ecosystem Genomics Graduate Interdisciplinary Program, University of Arizona, Tucson, AZ 85721, USA
[11]These authors contributed equally
[12]Lead contact
*Correspondence: arnold@ag.arizona.edu
https://doi.org/10.1016/j.cub.2024.01.063

SUMMARY

Understanding how symbiotic associations differ across environmental gradients is key to predicting the fate of symbioses as environments change, and it is vital for detecting global reservoirs of symbiont biodiversity in a changing world.[1–3] However, sampling of symbiotic partners at the full-biome scale is difficult and rare. As Earth's largest terrestrial biome, boreal forests influence carbon dynamics and climate regulation at a planetary scale. Plants and lichens in this biome host the highest known phylogenetic diversity of fungal endophytes, which occur within healthy photosynthetic tissues and can influence hosts' resilience to stress.[4,5] We examined how communities of endophytes are structured across the climate gradient of the boreal biome, focusing on the dominant plant and lichen species occurring across the entire south-to-north span of the boreal zone in eastern North America. Although often invoked for understanding the distribution of biodiversity, neither a latitudinal gradient nor mid-domain effect[5–7] can explain variation in endophyte diversity at this trans-biome scale. Instead, analyses considering shifts in forest characteristics, *Picea* biomass and age, and nutrients in host tissues from 46° to 58° N reveal strong and distinctive signatures of climate in defining endophyte assemblages in each host lineage. Host breadth of endophytes varies with climate factors, and biodiversity hotspots can be identified at plant-community transitions across the boreal zone at a global scale. Placed against a backdrop of global circumboreal sampling,[4] our study reveals the sensitivity of endophytic fungi, their reservoirs of biodiversity, and their important symbiotic associations, to climate.

RESULTS AND DISCUSSION

To evaluate how fungal symbiont communities shift across the environmental gradients that span an entire biome, we collected fresh photosynthetic material from common plants and lichens that occur across the south-to-north (SN) extent of the boreal biome in eastern North America (Figure 1): *Picea mariana* (Mill.) Britton, Sterns & Poggenb. (black spruce), *Pleurozium schreberi* (Brid.) Mitt (red-stemmed feather moss), and *Cladonia rangiferina* (L.) Nyl. (gray reindeer lichen). Each is abundant throughout the boreal zone of eastern North America, where *Pi. mariana* represents the dominant tree species (hereafter, *Picea*), *Pl. schreberi* is the most common moss (hereafter, *Pleurozium*), and *C. rangiferina* (hereafter, *Cladonia*) is the most common terricolous lichen found across the breadth of the boreal belt (Figure 1).[8] Collection sites (SN1–SN9) represented nine points along a biome-spanning latitudinal gradient (Figure 1A). Each site consisted of three subsites, and each subsite contained three microsites, for a total of 81 collection localities. Our sampling spanned 1,246 km in geographic distance, 11.2° of latitude, 9.6°C in mean annual temperature (MAT), 600 mm in mean annual precipitation (MAP), and vegetation types ranging from the northern edge of mixed hardwood forests—where birch and fir stands represent the southern extent of the boreal

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

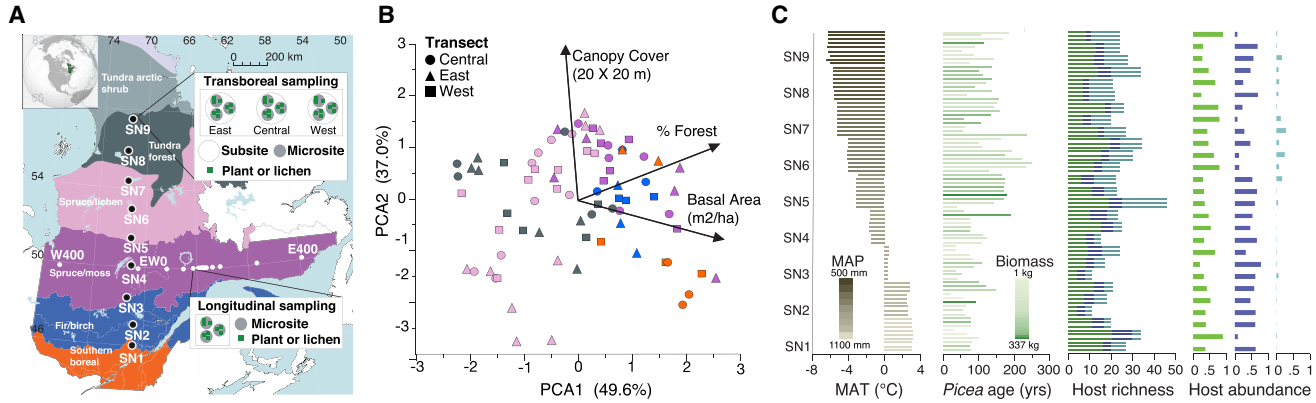

Figure 1. Sampling the diversity and composition of fungal endophyte communities at a trans-biome scale

(A) Québec, Canada: inset, position globally; colored map, vegetation types and sampling sites along the south-north (SN) transect. Each SN site (black circles, SN1–SN9) consists of three subsites (east, central, and west), each of which contains three microsites. We collected each of the three host taxa in each microsite. To consider spatial structure independent of climate, we conducted longitudinal sampling (white circles, east-west [EW] transect, E400-W400), collecting the 3 focal hosts in each of 3 microsites in 15 sites (white dots). For map source and license details, see STAR Methods.

(B) Forest characteristics measured in 20 × 20 m plots at each microsite, including canopy cover, percent of forested area, and basal area (STAR Methods); colors match (A).

(C) Climate and biotic data for SN1–SN9, including mean annual temperature (MAT, x axis) and mean annual precipitation (MAP, darkness of bars); age of sampled *Picea* individuals (x axis) and their estimated biomass (colors; STAR Methods); stacked bars indicating richness of vascular plants (green), non-vascular plants (dark blue), and lichens (teal); and relative abundance of focal hosts (*Picea*, green; *Pleurozium*, dark blue; *Cladonia*, teal) (see also Table S1).

See also Tables S2 and S3.

zone—to the southern edge of the Arctic biome (trans-biome sampling; Figure 1A). In each microsite, we measured forest cover, canopy cover, basal area, the diameter at breast height and age of individual *Picea* trees, the richness of plant and lichen communities, the relative abundance of focal host taxa, and carbon and nitrogen content of host tissue (Figures 1B and 1C; Table S1). We then characterized endophyte communities in surface-sterilized tissues of each species (leaves of *Picea*, phyllids of *Pleurozium*, or thalli of *Cladonia*) via barcode amplicon sequencing on the Illumina platform. We used both negative controls (extraction and PCR blanks) and positive controls (mock communities containing a diversity of fungal strains, prepared with equimolar and variable DNA concentrations) (STAR Methods). Rarefaction showed sufficient sampling depth across all sites for each host taxon, providing a basis for robust ecological analyses (Figure S1).

The latitudinal gradient of biodiversity often is invoked in defining landscape-scale distributions of species richness.[5,6] However, when all three host species were considered simultaneously, our survey across the SN scope of the boreal biome revealed that neither a simple latitudinal gradient (whereby diversity decreases with increasing latitude[5,6]) nor mid-domain effect (whereby diversity peaks in the midpoint of a geographic range[7]) could explain the distribution of species richness or phylogenetic diversity of endophytic fungi (Figures 2 and S2). Variation in endophyte richness could not be explained simply by host age or biomass at the scope of this study: multiple regression analyses of endophyte richness in *Picea*, for which we measured tree age by tree-ring analysis and estimated biomass via the allometric equation for that species, revealed the significance of climate factors (defined as the first principal component based on MAT and MAP; STAR Methods) in defining endophyte richness ($p = 0.01$), while neither age nor biomass of sampled individuals was significant ($p = 0.31$ and 0.21, respectively). Similarly, variation in endophyte richness at the trans-biome scale could not be explained simply by nutrient status of host tissue (i.e., C, N, or C:N ratio) or by forest characteristics (e.g., richness of host communities or vegetation factors such as the percent of forest to non-forest in the area, canopy cover, and basal area read of standing dead wood, Figure 1; Tables S1 and S2). Instead, we found that each host species had distinctive distributions of endophyte richness (Figures S1 and S2) and phylogenetic diversity, taxonomic composition, and community composition (Figure 2) across the SN extent of the boreal belt.

We predicted that these distinctive distributions of richness, phylogenetic diversity, and composition represented distinctive sensitivity of the symbioses between each host species and its endophytes to climate. A previous study at the circumboreal scale found that endophytic fungal communities are structured primarily by their hosts rather than climate differences.[4] However, that study focused on individual sites in Eurasia and North America and could not address regional climate gradients. In the present study, sampling the same host species in geographically proximate sites from the southern to northern extent of the boreal biome (spanning ca. 1,250 km), and from east to west (EW) at approximately the same latitude (also spanning ca. 1,250 km), provided the basis to identify the relevance of climate factors versus geographic distance alone, without the challenge of considering intercontinental or biogeographic differences in endophyte assemblages. Our observation of distinctive assemblages of endophytes in each host taxon, and host-specific differences in endophyte richness and phylogenetic diversity (Figures 2 and S1), led us to analyze data for each host species separately when testing our prediction that endophyte communities are structured by climate across the boreal biome.

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

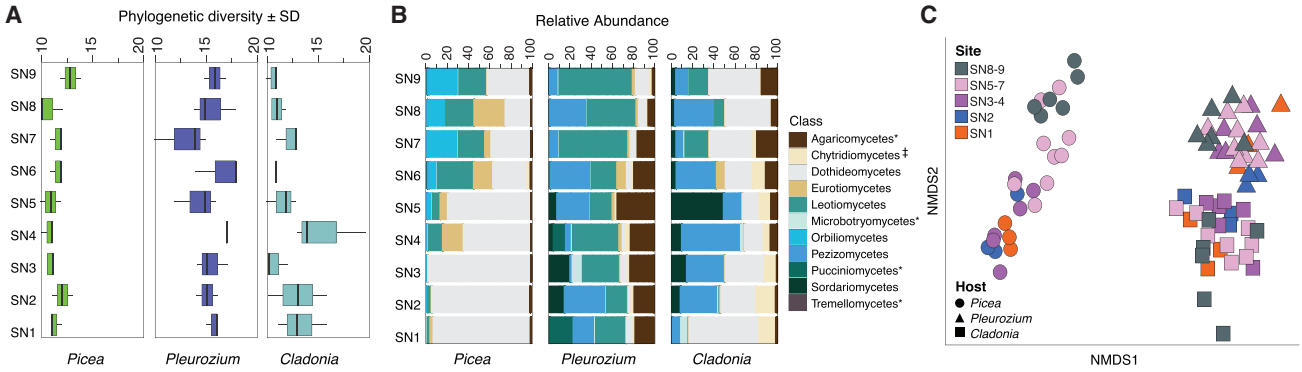

Figure 2. Differences in phylogenetic diversity of endophytic fungi, relative abundance of major endophytic taxa, and fungal community composition within and among focal hosts across the south-to-north extent of the boreal biome

(A) The scale and distribution of phylogenetic diversity of endophytes (number of fungal classes, including all classes shown in B, and additional classes that occurred at lower abundances; see publicly available data for this article) did not follow a simple latitudinal gradient for any host and differed among hosts.

(B) Phylogenetic composition of endophyte communities differed for each host across the gradient and among host taxa within sites. Summary data for all hosts in (A) and (B) are in Figure S2.

(C) Endophyte communities differed in composition among the three host taxa (NMDS; PERMANOVA, $R^2 = 0.27$; $p < 0.001$) and among sites in a host-specific manner. Colors match vegetation types in Figure 1A. Each sample was subsampled to 17,672 reads; individuals with <5,000 reads and OTU with <25 reads were excluded. Stress = 0.15.

See also Figure S2 and Tables S1, S3, and S4.

We examined the relationship of endophyte assemblages to climate for each locality studied here, and for all localities in which endophytes of these host species were sampled previously in circumboreal surveys, which allowed us to decouple latitude, forest characteristics, and climate[4] (Figure 1). In our analyses, climate represented MAT and MAP together as a single eigenvector based on a principal-component analysis (STAR Methods; this eigenvector represented >97% of climate variation across the SN gradient). In examining endophyte richness of each host species, we considered both linear and quadratic fits to permit species-specific mid-domain effects or other non-linear relationships,[7] with a focus on the SN gradient. We then examined how the dominant endophyte taxa in each host species varied in prevalence at the whole-biome scale, considering both our trans-biome sampling and previous sampling of the global circumboreal belt.[4]

Consistent with our prediction, differences in endophyte community composition for each host in the trans-biome sampling were best explained by major climatic factors captured by MAP and MAT (i.e., 12%–35% of variation in compositional shifts after accounting for inter-site distances; Table S3). However, increasing differences in climate among sites in the trans-biome survey occurred in parallel with increasing geographic distances between sites (Table S3). Therefore, to infer a direct role of climate factors in driving endophyte community structure, we considered alternative hypotheses that could explain endophyte distributions. For example, it is plausible that dispersal limitation at a landscape scale could drive the patterns we observed. We explored this in four ways.

First, if endophyte community structure is shaped primarily by dispersal limitation rather than climate, turnover in endophyte communities would increase with increasing distance among sites. However, we did not detect a simple, positive relationship between inter-site distance and turnover (estimated as beta diversity; Figure 3A). Instead, beta diversity of endophytes differed across the gradient, generally increasing among sites at the highest latitudes compared with lower latitudes (Figure 3A). This pattern was driven primarily by endophytes of two host taxa (*Picea* and *Cladonia*; Figure 3A). The strongest structuring of endophyte communities was observed for *Picea*, which demonstrated an especially notable transition in composition at the shift from spruce-moss to spruce-lichen communities from SN4 to SN5 (Figures 1A,3A, and 3B).

Consistent with this observation, latent Dirichlet allocation (LDA),[9] which can detect gradual shifts in communities, predicted a shift in the relative abundance of distinct endophyte communities for each host between SN4 and SN5 (Figure 3B). These sites correspond roughly to the latitude at which soils shift between discontinuous and continuous permafrost and forests transition to lichen-dominated understories[10] (Figure 1A). Permafrost impacts the depth of plant roots, and colder soil temperatures typically limit soil nutrient uptake and turnover, resulting in low plant productivity. This transition also corresponds to a marked reduction in the number of growing degree days in summer, which is hypothesized to impact tree productivity[11] and thus can influence the amount of carbon and nutrients available to fungal symbionts of leaves (Table S1).

Second, there is strong evidence that fungal endophytes are not dispersal-limited within sites,[4] but it is plausible that even airborne propagules could be dispersal-limited at the trans-biome scale despite the occurrence of hosts across all sites (Figure 1). If endophytes were dispersal-limited between but not within sites, we would expect that (1) endophyte communities in hosts located near to one another would be more similar to one another than to those in the same host species in other sites. Moreover, under such a scenario we would expect that (2) endophyte host use would not vary across the SN gradient: that is, beta diversity between hosts would be consistent across all sites. We examined these predictions through the lens of beta -diversity, first evaluating turnover in endophyte communities

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

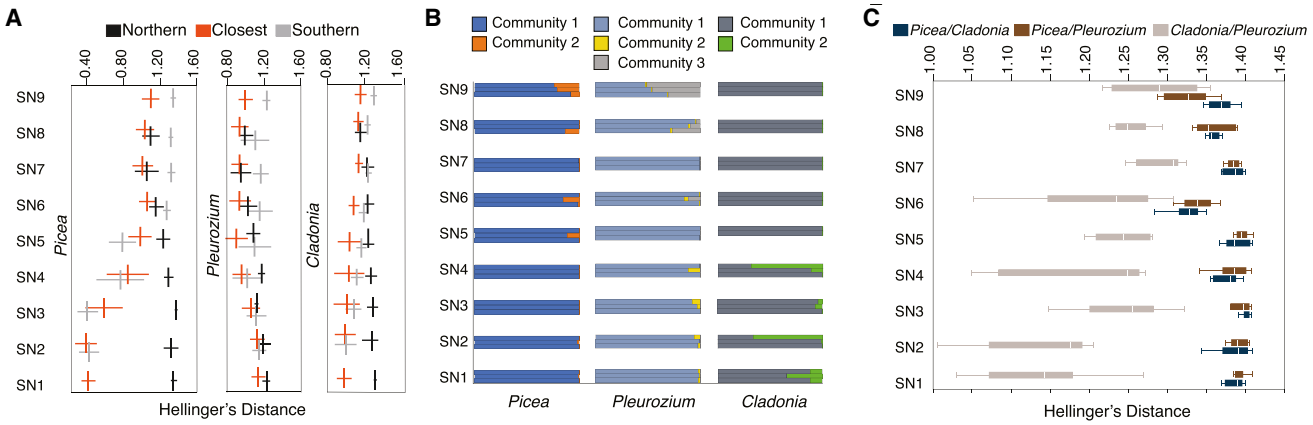

Figure 3. Endophyte community structure shifts at a bioclimatic transition zone linked to a change to lichen-dominated understories

(A) Endophyte communities in each SN site were compared with those in the next closest SN sites, those in the southernmost site (SN1), and those in the northernmost site (SN9). Beta diversity was calculated as Hellinger's distance. Sites north of SN5 tended to harbor endophyte communities that were more similar to the northernmost site (SN9) than to the southernmost site (SN1), and sites south of SN5 tended to harbor endophytes that were more similar to the southernmost site (SN1) than the northernmost site (SN9). Data for all hosts combined are shown in Figure S2.

(B) Latent Dirichlet allocation[9] (LDA) predicted at least two distinct communities ("components") per host genus across the SN gradient, with a transition where forests shift to lichen-dominated understories (near SN5). Each horizontal bar includes three rows that correspond to replicates (i.e., east, west, and central subsites); colors indicate the relative abundance of each component community per host genus.

(C) Pairwise dissimilarity of endophyte communities among host taxa shifts from SN1–SN9 ($p < 0.0001$ in all comparisons).

See also Figure S2 and Tables S1 and S3.

in the same host species in different sites and then evaluating changes in the dissimilarity of endophyte communities among host species at sites across the SN gradient (Figure 3).

In contrast to these predictions, we found that endophyte communities differed both among host species overall (Figure 2C) and between the same host species in different sites (Figures 2C and 3A). Moreover, the degree of turnover in endophyte communities between host taxa varied markedly across the gradient (Figure 3C; Table S4). Notably, communities of endophytes became more distinct between mosses and lichens at the colder, northern latitudes relative to the warmer, southern latitudes (Figure 3C). These lines of evidence underscore the relevance of climate in shaping endophyte communities at the trans-biome scale, and they are coherent with previous sampling along larger latitudinal gradients that predicted increasing host specificity at higher latitudes.[5]

Third, to provide insight into endophyte community shifts over geographic space with less marked differences in climate, we concurrently sampled hosts in 15 locations along a ca. 1,250 km transect in a single latitudinal band and vegetation type (spruce-moss forests; EW transect: 126 host collections; Figure 1A). The EW transect comprised <20% of the variation in MAT and <51% of the variation in MAP observed in the trans-biome sample (Table S1). Low turnover in endophyte community composition was observed along the EW transect (Table S4). Compositional changes in endophyte communities along the EW transect were not related to geographic distance between sites or the relatively small differences in climate among sites (Tables S3 and S4).

Finally, we placed our trans-biome results into a global circumboreal context, considering endophytes that were detected with the same methods in Eurasia and additional sites in North America[4] (Figure 4). We found that phylogenetic diversity of endophytes varied positively with climate dissimilarity across boreal forests worldwide. Endophyte community turnover (beta diversity) was high when climate conditions changed markedly among even relatively proximate sites, such as those along the SN gradient, consistent with climate filtering as a driver of community structure. This interpretation was supported by network analyses of the trans-biome dataset, which revealed that endophyte communities in each host differ among sites due to the high turnover in endophyte species (mean turnover = 0.88; Table S4).

Together, our data provide strong evidence that endophyte richness at the whole-biome scale varied with climate and, moreover, that it did so in a host-specific manner (Figure 4; Table S5). For example, endophytes of *Picea* generally decreased in richness in warmer and wetter climates (Figure 4A), where communities were largely dominated by a putatively novel species of unknown function that is closely related to the conifer needle pathogen *Nothophaeocryptopus gaeumannii* (Capnodiales, Dothideomycetes) (Figure S2). Fungal endophytes of *Pleurozium* peaked in richness in both colder and drier and warmer and wetter extremes (Figure 4B). Endophytes of *Cladonia* increased in richness with warmer, wetter conditions (Figure 4C), consistent with the high species richness of endophytes observed in this genus and other lichens in temperate forests.[5,12]

Overall, endophyte richness was very high across the trans-biome gradient because few fungal endophyte species were shared among sites. Instead, climate-specific assemblages were found with each host species at sites along the gradient (Tables S4 and S5). Network analyses found that less diverse endophyte communities did not represent subsets of the species at more diverse sites (i.e., nestedness was low; mean = 0.03; Table S4). We anticipated that the sensitivity of endophytes to climate would be detectable in terms of observed endophyte

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

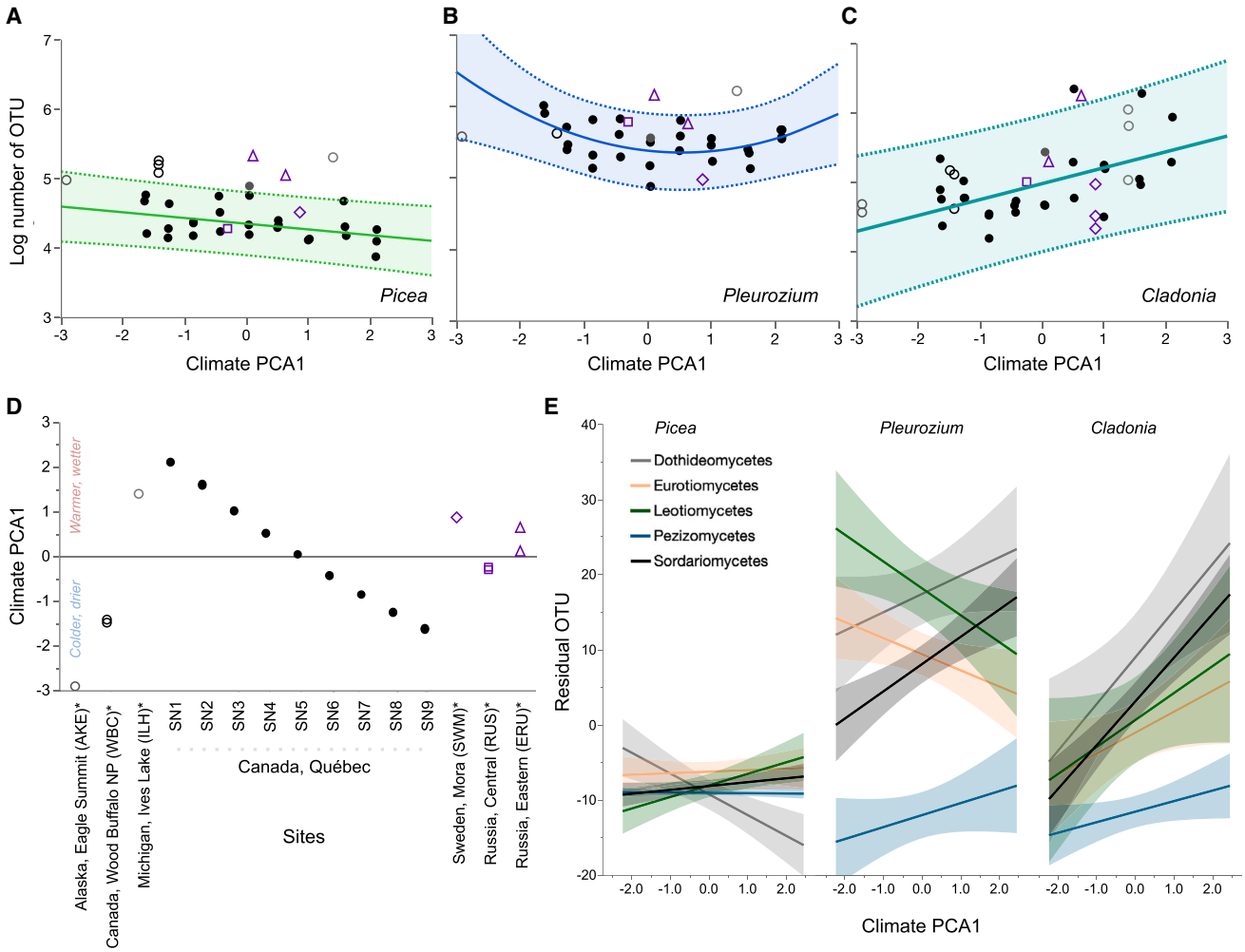

Figure 4. Relationship of endophyte richness and climate differs among hosts and fungal clades and identifies endophyte biodiversity hotspots at a global scale

(A–E) Richness of endophytes in *Picea*, *Pleurozium*, and *Cladonia* (A–C), SN1–SN9 (solid circles), as a function of climate (higher values = warmer and wetter). Data were rarefied as in Figure 2. Bands show 95% confidence intervals for transboreal data. Open circles indicate endophytes of congeneric hosts from North America (black) and Eurasia (purple), as shown in (D). Climate data for sites from the transboreal (solid black circles) and circumboreal sites (open symbols, U'Ren et al.[4]) (D). Decomposition of (A)–(C) reveals that richness in the most common fungal classes, presented as the residual richness after adjustment by the square root of the read number, varies with climate and host (E). For statistical analyses, see Table S5.
See also Tables S1 and S2.

species (Figures 4A–4C) and also the evolutionary history of those endophyte-containing lineages. This evolutionary signal would be expected to be especially strong if, as observed here, endophytes affiliate preferentially with particular lineages with which they may have established climate-specific symbioses. We therefore tested the prediction that the richness of endophytes in the five most common classes of Pezizomycotina also varied with climate in a host-specific manner.

As anticipated, trans-biome sampling showed that the richness of these classes differs both as a function of climate and in the context of symbioses with different host taxa. For example, the richness of Dothideomycetes in *Picea* diminished toward warmer and wetter conditions, becoming dominated by a single taxon (Figures 4A, 4D, and 4E; Table S5). In *Pleurozium*, the richness of Dothideomycetes, Eurotiomycetes, and Leotiomycetes increased at the ends of the gradient (Figures 4B, 4D, and 4E). Overall, the trend in all fungal classes occurring as endophytes of *Cladonia* (Figures 4C–4E), which has a range that extends far into the Arctic, was to decrease in richness under colder and drier conditions. This pattern was most pronounced in classes known for their high species richness in temperate forests and other biomes to the south (i.e., Sordariomycetes and Dothideomycetes).[5,12] Such results are consistent with the expectation of fewer and relatively more host-specific endophytes in relatively extremophilic lichens as they become one of the predominant life forms at the transition to tundra.[5]

Our results complement and extend previous circumboreal sampling by showing that marked turnover can occur across contiguous locations that differ in climate (Figure S2). Trans-biome sampling captured representative phylogenetic diversity

of endophytes when compared against larger-scale circumboreal sampling (Figure S2), although Pezizomycetes were more prevalent in *Pleurozium* and *Cladonia* in Québec relative to sites across Eurasia and other regions of North America. Sampling across the boreal biome in Québec captured 33%–47% of the boreal endophyte diversity detected at a global circumboreal scale by U'Ren et al.[4] (Figure S2), highlighting the influence of climate-driven turnover in endophyte communities at local and regional scales.

The climate sensitivity of endophytes suggests that climate-informed data may be used to identify global hotspots of endophyte diversity. For example, when placed against a global backdrop of climate for sites surveyed at a circumboreal scale[4] (Figure 4D), richness values above the expected 95% confidence interval based on climate in the trans-biome surveys identified notable biodiversity hotspots for endophytes at locations across boreal forests worldwide (Figures 4A–4C). Among the sites we surveyed, these include boreal forests in Eastern Russia (endophytes of all three host genera), Northern Michigan, USA (endophytes of *Picea* and *Pleurozium*), and Northern Alberta, Canada (endophytes of *Picea*) (Figures 4A–4C). Together, these sites encompass a relatively wide range of climate conditions (Figure 4D), but they share a proximity to other plant communities and represent relatively heterogeneous landscapes with high local turnover in plant species over small geographic distances. In contrast, the sites across Québec and the other boreal sites studied previously[4] were typically located within large areas of contiguous forest with relatively consistent plant and lichen communities.

At a global scale, boreal forests represent ca. 30% of forest cover, have disproportionate effects on global climate and hydrology, and are massively threatened by climate change.[13–18] As ancient and foundational symbionts that preceded the evolution of mycorrhizal fungi[19] and exceed the diversity of all other guilds of plant-symbiotic fungi,[20] endophytes are central to responses of their hosts to climate stress over ecological and evolutionary time.[19–25] In experimental settings, endophytes are sensitive to climate shifts, and their functional roles can be altered as climate conditions change.[25,26] Our results show that across large environmental gradients, the diversity and composition of these symbiont communities is driven by climate but with variation among hosts, fungal lineages, and their associations. Thus, differences in endophyte communities are not easily characterized by geographic distances, latitude, or climate gradients alone, instead representing "bespoke symbioses," as differentially sensitive to climate factors as their hosts may be.[27,28] In a rapidly changing environment, the dual sensitivity of horizontally transferred symbionts and their hosts to climate may frame cascading losses of biodiversity at a local to global scale[28]: the iconic plants and lichens that define the rapidly changing boreal biome may diminish in resilience, not only through the direct effects of climate change but also due to losses of their climate-sensitive endophytic partners.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.01.063.

ACKNOWLEDGMENTS

We thank K. Arendt and E. Lefèvre for exceptional assistance in the field; T. Gleason and J. Riddle for laboratory assistance; R.J. Steidl for helpful discussions; K. Youens-Clark and T. O'Connor for computational assistance; A. Bruneau and C. Lutzoni for logistical assistance in Québec; and the University of Arizona Environmental Isotope Laboratory, the University of Arizona Genetics Core, and the University of Idaho IBEST Genomics Core for technical assistance. We especially thank the staff of Air Saguenay, particularly pilot Jacques Bérubé, for facilitating all aspects of our field work in Québec. S.O. was supported by the Funai Overseas Scholarship and by the School of Plant Sciences and College of Agriculture and Life Sciences at the University of Arizona. N.B.Z. was supported by the Gordon and Betty Moore Foundation through grant GBMF 2550.03 to the Life Sciences Research Foundation. This study was funded by the National Science Foundation Dimensions of Biodiversity program (A.E.A., DEB-1045766; F.L., DEB-1046065; I.C., DEB-1046167; and G.M., DEB-1045608). Data collection performed by the IBEST Genomics Resources Core at the University of Idaho was supported in part by the NIH (COBRE grant P30GM103324).

AUTHOR CONTRIBUTIONS

Conceptualization, F.L., A.E.A., and J.M., together with I.C. and G.M.; fieldwork, A.E.A., F.L., J.M., and B.B.; data collection, J.M.U., S.O., A.E.A., V.T., J.M., F.L., and B.B.; analysis development, J.M.U., S.O., and A.E.A.; data

analysis, J.M.U., S.O., D.V., and A.E.A.; writing – original draft, J.M.U., S.O., and A.E.A.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 29, 2023
Revised: December 3, 2023
Accepted: January 25, 2024
Published: February 16, 2024

REFERENCES

1. Arnold, A.E., Mejía, L.C., Kyllo, D., Rojas, E.I., Maynard, Z., Robbins, N., and Herre, E.A. (2003). Fungal endophytes limit pathogen damage in a tropical tree. Proc. Natl. Acad. Sci. USA *100*, 15649–15654.
2. Rodriguez, R.J., Henson, J., Van Volkenburgh, E., Hoy, M., Wright, L., Beckwith, F., Kim, Y.O., and Redman, R.S. (2008). Stress tolerance in plants via habitat-adapted symbiosis. ISME J. *2*, 404–416.
3. Busby, P.E., Soman, C., Wagner, M.R., Friesen, M.L., Kremer, J., Bennett, A., Morsy, M., Eisen, J.A., Leach, J.E., and Dangl, J.L. (2017). Research priorities for harnessing plant microbiomes in sustainable agriculture. PLoS Biol. *15*, e2001793.
4. U'Ren, J.M., Lutzoni, F., Miadlikowska, J., Zimmerman, N.B., Carbone, I., May, G., and Arnold, A.E. (2019). Host availability drives distributions of fungal endophytes in the imperilled boreal realm. Nat. Ecol. Evol. *3*, 1430–1437.
5. Arnold, A.E., and Lutzoni, F. (2007). Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? Ecology *88*, 541–549.
6. Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. Am. Nat. *163*, 192–211.
7. Colwell, R.K., and Lees, D.C. (2000). The mid-domain effect: geometric constraints on the geography of species richness. Trends Ecol. Evol. *15*, 70–76.
8. McLaren, J.B., and Holguin, J. (2023). Boreal forest ecosystems. In Encyclopedia of Biodiversity, S.M. Scheiner, ed. (Academic Press), pp. 350–362.
9. Valle, D., Baiser, B., Woodall, C.W., and Chazdon, R. (2014). Decomposing biodiversity data using the latent dirichlet allocation model, a probabilistic multivariate statistical method. Ecol. Lett. *17*, 1591–1601.
10. Payette, S. (1992). Fire as a controlling process in the North American boreal forest. In A Systems Analysis of the Global Boreal Forest, H.H. Shugart, R. Leemans, and G.B. Bonan, eds. (Cambridge University Press), pp. 144–169.
11. Nicault, A., Boucher, E., Tapsoba, D., Arseneault, D., Berninger, F., Bégin, C., DesGranges, J.L., Guiot, J., Marion, J., Wicha, S., and Bégin, Y. (2015). Spatial analysis of black spruce (*Picea mariana* (Mill.) B.S.P.) radial growth response to climate in northern Québec – Labrador Peninsula, Canada. Can. J. For. Res. *45*, 343–352.
12. U'Ren, J.M., Lutzoni, F., Miadlikowska, J., Laetsch, A.D., and Arnold, A.E. (2012). Host and geographic structure of endophytic and endolichenic fungi at a continental scale. Am. J. Bot. *99*, 898–914.
13. Frelich, L.E. (2013). Boreal Biome. Oxford Bibliographies Online Datasets. https://doi.org/10.1093/obo/9780199830060-0085.
14. Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A.Z., and Schepaschenko, D.G. (2015). Boreal forest health and global change. Science *349*, 819–822.
15. Peng, C., Ma, Z., Lei, X., Zhu, Q., Chen, H., Wang, W., Liu, S., Li, W., Fang, X., and Zhou, X. (2011). A drought-induced pervasive increase in tree mortality across Canada's boreal forests. Nature Clim. Change *1*, 467–471.
16. Juday, G.P., Alix, C., and Grant, T.A. (2015). Spatial coherence and change of opposite white spruce temperature sensitivities on floodplains in Alaska confirms early-stage boreal biome shift. For. Ecol. Manag. *350*, 46–61.
17. Hogg, E.H., Michaelian, M., Hook, T.I., and Undershultz, M.E. (2017). Recent climatic drying leads to age-independent growth reductions of white spruce stands in western Canada. Glob. Chang. Biol. *23*, 5297–5308.
18. Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.B., Pe'er, G., Singer, A., Bridle, J.R., Crozier, L.G., De Meester, L., Godsoe, W., et al. (2016). Improving the forecast for biodiversity under climate change. Science *353*, aad8466.
19. Lutzoni, F., Nowak, M.D., Alfaro, M.E., Reeb, V., Miadlikowska, J., Krug, M., Arnold, A.E., Lewis, L.A., Swofford, D.L., Hibbett, D., et al. (2018). Contemporaneous radiations of fungi and plants linked to symbiosis. Nat. Commun. *9*, 5451.
20. Blackwell, M. (2011). The fungi: 1, 2, 3... 5.1 million species?†. Am. J. Bot. *98*, 426–438.
21. U'Ren, J.M., Lutzoni, F., Miadlikowska, J., and Arnold, A.E. (2010). Community analysis reveals close affinities between endophytic and endolichenic fungi in mosses and lichens. Microb. Ecol. *60*, 340–353.
22. Rodriguez, R., and Redman, R. (2008). More than 400 million years of evolution and some plants still can't make it on their own: plant stress tolerance via fungal symbiosis. J. Exp. Bot. *59*, 1109–1114.
23. Colón-Carrión, N., Lozada-Troche, C.A., and Arnold, A.E. (2022). Communities of endophytic fungi in a Puerto Rican rainforest vary along a gradient of disturbance due to Hurricane Maria. Ecol. Evol. *12*, e9618.
24. Müller, D.B., Vogel, C., Bai, Y., and Vorholt, J.A. (2016). The plant microbiota: systems-level insights and perspectives. Annu. Rev. Genet. *50*, 211–234.
25. Giauque, H., and Hawkes, C.V. (2013). Climate affects symbiotic fungal endophyte diversity and performance. Am. J. Bot. *100*, 1435–1444.
26. Treseder, K.K., Marusenko, Y., Romero-Olivares, A.L., and Maltz, M.R. (2016). Experimental warming alters potential function of the fungal community in boreal forest. Glob. Chang. Biol. *22*, 3395–3404.
27. Zook, D.P. (2001). Prioritizing symbiosis to sustain biodiversity: are symbionts keystone species? In Symbiosis: Mechanisms and Model Systems, J. Seckbach, ed. (Springer Netherlands), pp. 3–12.
28. Trevelline, B.K., Fontaine, S.S., Hartup, B.K., and Kohl, K.D. (2019). Conservation biology needs a microbial renaissance: a call for the consideration of host-associated microbiota in wildlife management practices. Proc. Biol. Sci. *286*, 20182448.
29. U'Ren, J.M., Lutzoni, F., Miadlikowska, J., Zimmerman, N.B., Carbone, I., May, G., and Arnold, A.E. (2019). Host availability drives distributions of fungal endophytes in the imperiled boreal realm. (Figshare). https://doi.org/10.6084/m9.figshare.c.4327772.
30. Videira, S.I.R., Groenewald, J.Z., Nakashima, C., Braun, U., Barreto, R.W., de Wit, P.J.G.M., and Crous, P.W. (2017). Mycosphaerellaceae - Chaos or clarity? Stud. Mycol. *87*, 257–421.
31. Brewer, P.W. (2014). Data management in dendroarchaeology using Tellervo. Radiocarbon *56*, S79–S83.
32. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.
33. Edgar, R.C., and Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics *31*, 3476–3482.
34. Huson, D.H., and Mitra, S. (2012). Introduction to the analysis of environmental sequences: metagenomics with MEGAN. Methods Mol. Biol. *856*, 415–429.
35. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.
36. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

37. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

38. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

39. R Core Team (2017). R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org.

40. Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M.H.H., Oksanen, M.J., and Suggests, M. (2018). vegan: Community ecology package. R package version 2.4–6. https://CRAN.R-project.org/package=vegan.

41. McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One *8*, e61217.

42. Nychka, D., Furrer, R., Paige, J., and Sain, S. (2015). fields: Tools for Spatial Data. https://doi.org/10.5065/D6W957CT.

43. Dormann, C.F., Gruber, B., and Fründ, J. (2008). Introducing the bipartite package: analysing ecological networks. Interaction *1*, 8–11. http://www.biom.uni-freiburg.de/Dateien/PDF/dormann2008rnews.pdf.

44. Baselga, A., and Orme, C.D.L. (2012). betapart: an R package for the study of beta diversity. Methods Ecol. Evol. *3*, 808–812.

45. Holmes, R.L., Adams, R., and Fritts, H.C. (1986). Quality control of cross-dating and measuring: a user's manual for program COFECHA. In Tree Ring Chronologies of Western North America: California, Eastern Oregon and Northern Great Basin (University of Arizona, Laboratory of Tree-Ring Research), pp. 41–49.

46. Gonzalez, M.A. (2001). Recent formation of arroyos in the Little Missouri Badlands of southwestern North Dakota. Geomorphology *38*, 63–84.

47. Lambert, M.-C., Ung, C.-H., and Raulier, F. (2005). Canadian national tree aboveground biomass equations. Can. J. For. Res. *35*, 1996–2018.

48. Higgins, K.L., Coley, P.D., Kursar, T.A., and Arnold, A.E. (2011). Culturing and direct PCR suggest prevalent host generalism among diverse fungal endophytes of tropical forest grasses. Mycologia *103*, 247–260.

49. U'Ren, J.M., Riddle, J.M., Monacell, J.T., Carbone, I., Miadlikowska, J., and Arnold, A.E. (2014). Tissue storage and primer selection influence pyrosequencing-based inferences of diversity and community composition of endolichenic and endophytic fungi. Mol. Ecol. Resour. *14*, 1032–1048.

50. U'Ren, J.M. (2016). DNA Extraction from Fungal Mycelium Using Extract-n-Amp. https://doi.org/10.17504/protocols.io.ga4bsgw.

51. Gardes, M., and Bruns, T.D. (1993). ITS primers with enhanced specificity for basidiomycetes–application to the identification of mycorrhizae and rusts. Mol. Ecol. *2*, 113–118.

52. Vilgalys, R., and Hester, M. (1990). Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several Cryptococcus species. J. Bacteriol. *172*, 4238–4246.

53. U'Ren, J.M., and Arnold, A.E. (2017). DNA Extraction Protocol for Plant and Lichen Tissues Stored in CTAB. https://doi.org/10.17504/protocols.io.fs8bnhw.

54. U'Ren, J.M., and Arnold, A.E. (2017). Illumina MiSeq Dual-Barcoded Two-Step PCR Amplicon Sequencing Protocol. https://doi.org/10.17504/protocols.io.fs9bnh6.

55. White, T.J., Bruns, T., Lee, S., and Taylor, J.W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In PCR – Protocols and Applications – A Laboratory Manual (Academic Press), pp. 315–322.

56. Daru, B.H., Bowman, E.A., Pfister, D.H., and Arnold, A.E. (2018). A novel proof of concept for capturing the diversity of endophytic fungi preserved in herbarium specimens. Philos. Trans. R. Soc. Lond. B Biol. Sci. *374*, 20170395.

57. Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., et al. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol. Evol. *4*, 914–919.

58. Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods *10*, 996–998.

59. Abarenkov, K., Henrik Nilsson, R., Larsson, K.H., Alexander, I.J., Eberhardt, U., Erland, S., Høiland, K., Kjøller, R., Larsson, E., Pennanen, T., et al. (2010). The UNITE database for molecular identification of fungi–recent updates and future perspectives. New Phytol. *186*, 281–285.

60. Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics *27*, 2194–2200.

61. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

62. Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. *73*, 5261–5267.

63. Edgar, R.C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. Preprint at bioRxiv. https://doi.org/10.1101/081257.

64. U'Ren, J.M., Dalling, J.W., Gallery, R.E., Maddison, D.R., Davis, E.C., Gibson, C.M., and Arnold, A.E. (2009). 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.

65. U'Ren, J.M., Miadlikowska, J., Zimmerman, N.B., Lutzoni, F., Stajich, J.E., and Arnold, A.E. (2016). Contributions of North American endophytes to the phylogeny, ecology, and taxonomy of Xylariaceae (Sordariomycetes, Ascomycota). Mol. Phylogenet. Evol. *98*, 210–232.

66. Edgar, R.C. (2018). UNCROSS2: identification of cross-talk in 16S rRNA OTU tables. Preprint at bioRxiv. https://doi.org/10.1101/400762.

67. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods *13*, 581–583.

68. Tedersoo, L., and Lindahl, B. (2016). Fungal identification biases in microbiome projects. Environ. Microbiol. Rep. *8*, 774–779.

69. Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., and Kauserud, H. (2010). ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol. *10*, 189.

70. Stone, J.K., Capitano, B.R., and Kerrigan, J.L. (2008). The histopathology of *Phaeocryptopus gaeumannii* on Douglas-fir needles. Mycologia *100*, 431–444.

71. Anderson Stewart, C.R., Doilom, M., and Taylor, J.E. (2019). Analysis of fungal endophytes in Scottish Sitka spruce plantations shows extensive infections, novel host partners and gives insights into origins. For. Pathol. *49*, e12471.

72. Nguyen, D., Boberg, J., Ihrmark, K., Stenström, E., and Stenlid, J. (2016). Do foliar fungal communities of Norway spruce shift along a tree species diversity gradient in mature European forests? Fungal Ecol. *23*, 97–108.

73. McMullin, D.R., Green, B.D., Prince, N.C., Tanney, J.B., and Miller, J.D. (2017). Natural products of *Picea* endophytes from the Acadian Forest. J. Nat. Prod. *80*, 1475–1483.

74. Crous, P.W., Schoch, C.L., Hyde, K.D., Wood, A.R., Gueidan, C., de Hoog, G.S., and Groenewald, J.Z. (2009). Phylogenetic lineages in the Capnodiales. Stud. Mycol. *64*, 17–47S7.

75. Weiss, S.J., Xu, Z., Amir, A., Peddada, S., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vazquez-Baeza, Y., Birmingham, A., and Knight, B. (2015). Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data. Preprint at PeerJ Preprints. https://dx.doi.org/10.7287/peerj.preprints.1157v1.

76. Zimmerman, N.B., and Vitousek, P.M. (2012). Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. Proc. Natl. Acad. Sci. USA *109*, 13022–13027.

77. Barge, E.G., Leopold, D.R., Peay, K.G., Newcombe, G., and Busby, P.E. (2019). Differentiating spatial from environmental effects on

foliar fungal communities of *Populus trichocarpa*. J. Biogeog. *46*, 2001–2011.

78. Borcard, D., and Legendre, P. (2012). Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. Ecology *93*, 1473–1481.
79. Borcard, D., and Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Modell. *153*, 51–68.
80. Dray, S., Legendre, P., and Peres-Neto, P.R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Modell. *196*, 483–493.
81. Legendre, P., and Anderson, M.J. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol. Monogr. *69*, 1–24.
82. Legendre, P., and Legendre, L.F.J. (2012). Numerical Ecology, Third Edition (Elsevier).
83. Blüthgen, N., Menzel, F., and Blüthgen, N. (2006). Measuring specialization in species interaction networks. BMC Ecol. *6*, 9.
84. Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. Glob. Ecol. Biogeogr. *19*, 134–143.
85. Borcard, D., Legendre, P., and Drapeau, P. (1992). Partialling out the spatial component of ecological variation. Ecology *73*, 1045–1055.

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Plant and lichen samples	This paper, Table S1, and U'Ren et al.[29]	https://doi.org/10.6084/m9.figshare.c.5005817; https://doi.org/10.6084/m9.figshare.c.4327772
Fungal endophyte cultures from surface-sterilized plants and lichens	This paper, Table S6, and U'Ren et al.[29]	https://doi.org/10.6084/m9.figshare.c.5005817; https://doi.org/10.6084/m9.figshare.c.4327772
Deposited data		
Metadata for sampling sites and hosts	This paper, Table S1, and and U'Ren et al.[29]	https://doi.org/10.6084/m9.figshare.c.5005817; https://doi.org/10.6084/m9.figshare.c.4327772
Raw data for transboreal sampling (SN transect), including raw Illumina ITS2 sequences of fungal endophytes	This paper, Table S1	BioProject PRJNA647873, SRA BioSamples SAMN15641885-SAMN15642127
Raw data for longitudinal sampling (EW transect), including raw ITS1 sequences of fungal endophytes	This paper, Table S1	BioProject PRJNA647873, SRA BioSamples SAMN39598376-SAMN39598441
ITS-LSU nrDNA Sanger sequences and metadata for endophytes isolated in culture	This paper, Table S6, and U'Ren et al.[29]	https://doi.org/10.6084/m9.figshare.c.5005817; https://doi.org/10.6084/m9.figshare.c.4327772
Raw data for circumboreal samples including raw Illumina ITS2 sequences of fungal endophytes	This paper, Table S1, and U'Ren et al.[29]	https://doi.org/10.6084/m9.figshare.c.5005817; https://doi.org/10.6084/m9.figshare.c.4327772; BioProject PRJNA514023, SRA BioSamples listed in Table S1
LSU nrDNA sequences for phylogenetic analyses	This paper, and Videira et al.[30]	[https://doi.org/10.1016/j.simyco.2017.09.003]
Oligonucleotides		
ITS1F: CTTGGTCATTTAGAGGAAGTAA	IDT DNA	N/A
ITS4: TCCTCCGCTTATTGATATGC	IDT DNA	N/A
LR3: GGTCCGTGTTTCAAGAC	IDT DNA	N/A
LSU1fd: GRATCAGGTAGGRATACCCG	IDT DNA	N/A
LR5: TCCTGAGGGAAACTTCG	IDT DNA	N/A
Software and algorithms		
Latent Dirichlet Allocation (LDA) algorithm	Valle et al.[9]	https://doi.org/10.1111/ele.12380
Tellervo development	Brewer[31]	http://www.tellervo.org/
USEARCH v10.0.240	Edgar,[32] Edgar and Flyvbjerg[33]	https://www.drive5.com/usearch/
MEGAN v5.11.3	Huson and Mitra[34]	https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/algorithms-in-bioinformatics/software/megan6/ (Most updated version)
QIIME v. 1.8	Caporaso et al.[35]	http://qiime.org/
MAFFT v7	Katoh and Standley[36]	https://mafft.cbrc.jp/alignment/software/
RAxML v8.2.12	Stamatakis[37]	https://cme.h-its.org/exelixis/web/software/raxml/
cutadapt 1.16	Martin[38]	https://cutadapt.readthedocs.io/en/v1.16/
JMP 12, 13	SAS Institute, Cary, NC, USA	https://www.jmp.com/en_us/home.html
R statistical environment (open-source software) v3.6.0	R Core Team[39]	https://cran.r-project.org
R package *vegan*	Oksanen et al.[40]	https://cran.r-project.org/web/packages/vegan/index.html
R package *phyloseq*	McMurdie and Holmes[41]	https://joey711.github.io/phyloseq/
R package *fields*	Nychka et al.[42]	https://cran.r-project.org/web/packages/fields/index.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R package *bipartite*	Dormann et al.[43]	https://cran.r-project.org/web/packages/bipartite/index.html
R package *betapart*	Baselga and Orme[44]	https://cran.r-project.org/web/packages/betapart/betapart.pdf

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, A. Elizabeth Arnold (arnold@ag.arizona.edu).

Materials availability

Cultures generated in this study were deposited as living vouchers at the Robert L. Gilbertson Mycological Herbarium, University of Arizona (Table S6). Host specimens were deposited at the Duke University Herbarium (Table S1) and at the University of Arizona.

Data and code availability

- Raw sequence data and metadata are deposited at DDBJ/EMBL/GenBank (sampling in Québec: BioProject PRJNA647873, SRA BioSamples SAMN15641885- SAMN15642127 and SAMN39598376-39598441; Circumboreal: BioProject PRJNA514023, SRA BioSamples listed in Table S1).
- All original code used in this study is publicly available in Figshare as of the date of publication. DOIs are listed in the key resources table.
- All other data are released with this article in the supplemental information.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Field collections

Field sampling on the trans-biome scale was conducted in the height of the growing season in July 2011 in the boreal biome of Québec, Canada (Figure 1A). With access by float plane, we sampled along a trans-biome latitudinal transect (south-north, SN), which consisted of nine major points (sites SN1-SN9; Figure 1A) spaced at ca. ~160 km intervals from the southern limit of the boreal biome to the northern edge at the tree limit (Figure 1A; Table S1). In total, sampling spanned 11.2° of latitude (46.76° - 57.96° N) and 1,246 km south-to-north (SN, Figure 1), encompassing sites that differ in mean annual temperature (MAT) by 9.6°C (range, 3.1°C to -6.5°C), mean annual precipitation (MAP) by 600 mm (range, 497 - 1097 mm), and vegetation structure and associated forest characteristics (Figure 1; Table S1). Each latitudinal point consisted of three subsites (East, Central, and West), each located ca. 3.2 km to the west (W) and east (E) of a central (C) subsite. Sampling in each of the three subsites (W, C, and E) was conducted in three replicate microsites (locations, M1-M3; ca. 20 m in diameter) located ~10-20 m apart (see Figure 1A), for a total of 81 sampling locations on the SN transect (Table S1).

Concurrently we sampled along a longitudinal transect (east-west, EW), located at roughly the latitude of SN4 (51.03°N), which spanned ca. 1250 km (18.4° of longitude) from the westernmost (W400: 50.28°N, 77.50°W) to the easternmost site (E400: 51.68°N, 59.15°W (Table S1). The 14 EW sites encompassed only 2.1° of latitude and as a result, climatic variation was limited among them (range of MAT, 1.8°C; range of MAP, 310 mm (Table S1). Instead of being evenly spaced as in the SN transect, sites along the EW transect ranged from 2.4 km to over 600 km from the central locality (i.e., EW0: 51.11°N, 68.52°W; located ~320 km to the east of SN4) to facilitate evaluation for distance decay in the context of relatively little environmental change over geographic distance. Each EW sampling location consisted of a single subsite, each with three microsites (Figure 1A).

Circumglobal sampling encompassed seven boreal forest sites in North America and Eurasia that together encompassed ca. 18.6° of latitude (46.85°–65.48°N) and 218° of longitude (135.96° to -145.42°), with a range of MAT of 12.6°C (-8.1°C to 4.5°C), and a range of MAP of 816 mm (281 - 1097 mm) (see U'Ren et al.[4,29] and Table S1).

Vegetation zone information and the base map modified for Figure 1 were obtained from the Ministère des Forêts, de la Faune et des Parcs Québec (accessed online, fall 2019: https://mffp.gouv.qc.ca/les-forets/amenagement-durable-forets/inventaire-ecoforestier/) and modified for use under a ShareAlike 4.0 International Creative Commons License.

For the present study, in each microsite we collected three small branches containing healthy needles of *Picea mariana* (Mill.) Britton, Sterns & Poggenb. (black spruce), one small mat (4–9 cm^2) of the moss *Pleurozium schreberi* (Brid.) Mitt. (red-stemmed feather moss), and one mature thallus of *Cladonia rangiferina* (L.) Nyl. (grey reindeer lichen) (Table S1). These host species are representative of the boreal biome in eastern Canada, with distributions spanning the entirety of the SN and EW transects. In addition, previous work

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

demonstrated these hosts harbor diverse and abundant endophytic communities.[4,5] The same genera were sampled with the same methods in U'Ren et al.[4] at a circumboreal scale. Sufficient material was collected for DNA extraction and endophyte isolations, chemical and isotopic analyses, and herbarium specimens for each collection. All lichen specimens were deposited at DUKE (Table S1) and material from plant collections has been archived at the University of Arizona. In total, we sampled 81 host collections per species along the SN gradient (243 total host collections) and 42 host collections per species along the EW transect (126 total host collections).

METHOD DETAILS

Characterization of environmental factors

In each microsite, we recorded details on the richness and composition of vascular and non-vascular plant and lichen communities (i.e., the total number of vascular plants, non-vascular, plants, and non-rock inhabiting lichens, as well as the relative abundance of each of the three focal hosts per sampling location), basal area forest cover (m^2/ha), percent forest to non-forest (estimated visually from the float plane), percent canopy cover, basal area read of standing dead wood (m^2/ha), and fire history in each microsite (Table S1). We observed no evidence of recent fire in any microsite (based on tree cores and observations of fire damage; i.e., charcoal, scarring, and related indicators), which was verified by interviews with forestry agents and forestry data (when available). Climate data for each sampling location were obtained from the WorldClim database (www.worldclim.org) at 30 arcsecond resolution. Metadata for each sampling location are provided in Table S1. Due to multicollinearity among climatic variables along the SN latitudinal gradient, we compared climate among sites using the eigenvector with the highest eigenvalue (i.e., first principal component) from the principal component analysis (PCA) of two variables (mean annual temperature, MAT; mean annual precipitation, MAP). The eigenvector (PCA1) explained 97.3% of the variation in climate at the trans-biome scale in terms of MAT and MAP.

Estimation of tree age and biomass

We estimated the height of each *Picea* individual sampled for endophytes, measured the diameter of those trees at breast height (DBH), and collected one tree core at the base of each individual. To estimate age of each *Picea* individual, we prepared tree cores by sanding until tree rings were clearly visible. After cores were cross-dated visually, we used the LINTAB measuring system (RinnTech) with Tellervo software[31] to measure tree ring width for further cross-dating. Cross-dating quality was checked with COFECHA across all tree cores from the same site.[45] We used a pith locator to reduce dating error for cores without piths.[46] Average growth rate was calculated as the number of tree rings divided by the total length of tree ring width. We used the DBH of each tree for the allometric equation established specifically for *Picea mariana* in Canada to estimate tree biomass.[47] Tree cores have been archived at the Trouet lab at the University of Arizona. We could not measure age or estimate biomass of *Pleurozium* or *Cladonia*, and thus analyses considering age and biomass of hosts refer only to *Picea*.

Tissue processing

Fresh tissues from each host collection were cut into 2 mm^2 segments, which were surface-sterilized following U'Ren et al.[4] Ninety-six surface-sterilized segments per host collection were chosen haphazardly and placed in CTAB buffer (1 M Tris HCl pH 8, 5 M NaCl, 0.5 M EDTA and 20 g CTAB) under sterile conditions.[48,49] CTAB tubes were stored at room temperature until shipped to the University of Arizona, where they were stored at -80 °C until DNA was extracted (see methods below). In two focal sites for each transect (SN5C and EW0) we haphazardly chose another 96 tissue segments from each host collection to isolate endophytes in culture (see below for isolation details) to complement culture-free data from the same tissues.[4] Portable laminar flow hoods were used for sterile processing at remote locations, and sterile methods were used for all tissue processing steps within 24-48 h after collection. Extra photosynthetic material from each host collection that was used for endophyte isolations was air dried at room temperature and ~10 g of tissue was ground to a fine powder using a sterile mortar and pestle in liquid nitrogen. Ground tissues from 45 host collections (representing the odd SN subsites and microsites in the central locations; i.e., SN1C, SN3C, SN5C, SN7C, SN9) were used for analyses of carbon and nitrogen (see Table S1, methods below).

Endophyte isolation, DNA extraction, amplification, and Sanger sequencing

Surface-sterilized tissue pieces (96 per host collection) were placed on 2% malt extract agar (MEA) under sterile conditions.[4] Emergent fungi were vouchered in sterile water and deposited at the Robert L. Gilbertson Mycological Herbarium at the University of Arizona (Table S6). Total genomic DNA was extracted directly from each fungal isolate following U'Ren.[50] 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 using the primer pair ITS1F/LR3[51,52] and sequenced bidirectionally at the University of Arizona Genetics Core. Sanger sequences were manually curated to ensure accuracy.[12] High quality Sanger sequence data were obtained for 442 endophytic isolates obtained from the three host taxa in the two focal sites (SN5 and EW0; accession numbers listed in Table S6).

DNA extraction, amplification, and Illumina sequencing

Total genomic DNA was extracted from host collections along the SN gradient using a modified protocol for the MoBio PowerPlant Pro DNA Isolation Kit (Qiagen, Germantown, MD).[53] The fungal ITS nrDNA locus was amplified for each of the 243 collections from the SN gradient via a two-step library preparation process following U'Ren and Arnold[54] with the primer pair ITS1F/ITS4.[51,55] PCR for

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

each sample was performed in triplicate and amplification was verified on a 2% agarose gel 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) at the University of Idaho IBEST Genomics Core. All SN collections were sequenced on a single Illumina run in parallel with a phylogenetically diverse mock community as a positive control and extraction blanks and PCR water controls as negative controls following U'Ren et al.[4] Dedicated workspaces and other precautions against cross-contamination were as described in U'Ren et al.[4] and Daru et al.[56]

Bioinformatics

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, personal communication), as well as sequences containing > 1 mismatches to the barcode and > 4 mismatches to primers, were removed. The remaining 6,867,452 reads corresponding to the ITS2 nrDNA region were trimmed for quality using a truncation length of 170 bp and a maximum error rate of 1.0 in USEARCH v10.0.240,[32,33] resulting in 3,187,113 high-quality sequences. To combine Sanger sequences from cultures with Illumina sequences for direct comparisons, we first used ITSx 1.0.7[57] to identify Sanger sequences that did not contain at least 50 bp of either ITS1 or ITS2 nrDNA for removal (n = 2). For the remaining Sanger sequences (Table S6), all bases after the conserved region at the start of LSU nrDNA were removed and sequences were trimmed to a length of 170 bp to match the length of Illumina sequences. Sanger and Illumina sequences were dereplicated in parallel and clusters represented by only one or two Illumina sequences (i.e., singletons or doubletons) were removed.

After these filtering steps, dereplicated sequences from both the culture-based and culture-free Illumina analyses were clustered into operational taxonomic units (OTU) at 95% sequence similarity with USEARCH,[32] which accurately clustered sequences in the mock community (see U'Ren et al.[4] for details of clustering verification). In addition to *de novo* chimera checking performed during clustering,[58] representative sequences for each OTU were subjected to reference-based chimera checking using the UNITE database[59] with UCHIME.[60] Raw Illumina reads and all Sanger reads were mapped back to sequences of chimera-checked OTU to construct an OTU table containing 3,555,004 reads and at least 2,852 OTU. Analyses of the phylogenetically diverse mock community confirmed the accuracy of bioinformatic methods for estimates of species boundaries and consistent read counts among replicate samples for Illumina sequencing[4] (see below).

Following clustering, a representative sequence from each OTU was queried with ITSx[57] and OTU lacking the ITS2 region were removed from downstream analyses. Sequences from the remaining OTU were queried against NCBI nr (but excluding all unidentified, environmental sequences) with BLASTn.[61] BLAST output was analyzed in MEGAN v. 5.11.3[34] with default parameters for lowest common ancestor (LCA). OTU representing the lichen mycobiont or plant hosts, sequences with no hits, and/or sequences not classified to Fungi were removed from subsequent analyses. The remaining OTU were queried against the UNITE fungal database[59] with the RDP Classifier[62] for taxonomic classification using a cutoff threshold of 80% confidence as implemented in QIIME v. 1.8.[35]

Comparison of zOTU to 95% OTU

Clustering of all sequences from Illumina MiSeq and Sanger sequencing at 95% sequence similarity resulted in 2,852 OTU, whereas analyses of the same data using UNOISE2[63] zOTU resulted in a 1.4-fold increase in richness (3,989 zOTU after removing singletons and doubletons). However, the majority (i.e., 53%) of 95% OTU were represented by a single zOTU (mean zOTU per 95% OTU = 2.24 ± 2.18; range 1 to 24 zOTU per OTU). Although zOTU may reveal true sequence variation, our analyses of the mock community revealed spurious zOTU compared to when OTU were clustered at 95% (see below). Therefore, subsequent analyses were performed with OTU resulting from UPARSE due to the lower rate of spurious errors.

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 molecular grade 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 that would indicate contamination. All PCR NTCs were pooled and sequenced along with other samples on an Illumina MiSeq. We used these negative controls to assess the potential for OTU in our dataset to represent laboratory contamination following U'Ren et al.[4] and Daru et al.[56]

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.[4,56] DNA from each taxon was amplified individually in PCR1, quantified with a Qubit fluorometer, 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 to eliminate any possible variation that could occur among sequencing runs.

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

Mock community data were used to validate the bioinformatic approach we used to 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.[64,65] Here, we clustered OTU in USEARCH with UPARSE[32,58] 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, OTU defined at 95% ITS2 nrDNA sequence similarity resulted in the most accurate estimates of species boundaries and richness. Representative sequences for OTU 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 limited the number of spurious OTU resulting from sequencing errors (i.e., each mock taxon was represented by a single OTU; see U'Ren et al.[49]). We also confirmed that our bioinformatic methods limited spurious OTU due to barcode or tag shifting (i.e., 'cross-talk[66]'). After careful examination, we identified three OTU 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), rather than using only presence/absence measures that give equal weight to low-abundant OTU 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[63] (i.e., zero radius OTU; zOTU) and DADA2.[67] Quality control and trimming in UNOISE2 followed methods for UPARSE (i.e., maximum error rate of one, truncation at 170 bp). 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% OTU; 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.[58] 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 zOTU/ASV resulting from sequencing errors.[49]

Primer choice and rationale for sequencing

Fungal-specific primers that amplify the ITS2 region for fungi while excluding plants are not presently available.[68] We therefore amplified the entire ITS nrDNA region with the forward primer ITS1F[51] (which results in the fewest reads for plants during *in silico* PCR[69]) and the reverse primer ITS4. Thus, forward NGS reads (i.e., R1) corresponded to the ITS1 nrDNA region and the reverse NGS reads (i.e., R2) corresponded to the ITS2 region. Analysis of R1 and R2 reads yielded similar results, but we analyzed the R2 reads for two reasons. First, analysis of R2 allows direct comparisons to previous studies that used the ITS2 nrDNA region (see U'Ren et al.[4]). Second, because Sanger sequences for cultures extended into the LSU nrDNA region, they could be trimmed to match the exact start and end positions of the NGS sequences, providing a basis for comparison of OTU generated by culture-based and NGS approaches. Sanger sequences were trimmed manually as part of the sequence validation and editing process and thus did not all encompass the exact start position of R1 as generated by NGS. 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.

Chemical analyses of photosynthetic tissues

Carbon and nitrogen content of host tissues, as well as their isotopic fractions (δ15N and δ13C), were measured on a continuous-flow gas-ratio mass spectrometer (Finnigan Delta PlusXL) coupled to an elemental analyzer (Costech) at the University of Arizona Isotope Laboratory (Tables S1 and S3). Samples were combusted in the elemental analyzer and values were standardized based on acetanilide for elemental concentration, NBS-22 and USGS-24 for δ13C, and IAEA-N-1 and IAEA-N-2 for δ15N. Repeated comparisons to an internal standard revealed precision greater than ± 0.10 for δ13C and ± 0.2 for δ15N (1s).

Characterization of dominant Capnodiales OTU in *Picea*

The ITS2 sequence for the dominant Capnodiales OTU (OTU4) in *Picea* at lower latitudes along the SN gradient has 100% BLASTn sequence identity to sequences of *Nothophaeocryptopus gaeumannii*, the causal agent of Swiss needle cast on *Pseudotsuga menziesii*.[70] Other studies using NGS also have sequenced this same OTU in species of *Picea* sp. in boreal forests (i.e., Norway spruce[71,72]), but have not paired culture-free with culture-based methods to provide additional information on the identity of short NGS sequences (but see McMullin et al.[73]). Culturing efforts at the trans-biome scale yielded 13 representatives from two individuals of *Picea* in SN5C and EW0 (Table S6). For four of these cultures, we PCR -amplified the LSU nrDNA region using the primers LSU1fd[74] and LR5.[52] Amplicons were Sanger sequenced bi-directionally at the University of Arizona Genetics Core. Sequences were processed and edited as described above for Sanger ITS Sequencing. New sequences were aligned with reference taxa from Videira et al.[30] with the profile alignment feature in MAFFT v7[36] and analyzed in RAxMLv8.2.12[37] with 1000 bootstrap replicates (Figure S2). Phylogenetic analyses indicate OTU4 represents a putatively novel species of Capnodiales with close affinity to the needle pathogen, *N. gaeumannii*[71,73] (Figure S2). The non-pathogenic interaction of OTU4 with *Picea* was further supported with quantitative real-time PCR (rtPCR), which revealed that samples dominated by OTU4 do not contain higher fungal biomass relative to other *Picea* samples along the SN transect.

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

Molecular analyses for EW transect

Total genomic DNA was extracted for host collections along the EW transect with the Qiagen Plant DNeasy kit (Qiagen, Germantown, MD) with modifications following U'Ren et al.[49] The fungal ITS nrDNA locus was amplified for each of the 82 *Picea* and *Pleurozium* collections from the EW transect with the primer pair ITS1F/ITS4.[51,55] Amplification was verified on a 2% agarose gel stained with SYBR. Final PCR products were quantified fluorometrically, 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 sequencing on an Ion Torrent personal sequencer at the Duke University Sequencing Facility. EW collections were sequenced on five IonTorrent runs. Positive controls on all runs included the following: (i) three clone sequences from *Pleurozium* sequenced individually; (ii) a mixture of the three clone sequences from *Pleurozium* pooled in equimolar amounts and sequenced; and (iii) replicate samples sequenced across each run. The accuracy of OTU clustering methods was determined using positive controls as described above. Ion Torrent ITS1 sequences from *Picea* and *Pleurozium* from the EW transect were trimmed for primers and adapters with cutadapt 1.16 with Python 3.6.2.[38] The remaining sequences from each run were subsequently trimmed for quality with a truncation length of 200 bp and a maximum error rate of 0.75 in USEARCH.[32,33] Quality filtered sequences from all runs were pooled into a single file and dereplicated, and singleton and doubleton clusters were discarded. The remaining dereplicated sequences were clustered into operational taxonomic units (OTU) at 95% sequence similarity with USEARCH and a representative sequence for each OTU was subjected to reference-based chimera checking via the UNITE[59] database with UCHIME.[60] Raw Ion Torrent reads were mapped back to sequences of chimera-checked OTU to construct an OTU table containing 1,639,281 reads and 543 OTU. We did not compare endophyte richness among samples analyzed by two different sequencing platforms (i.e., IonTorrent for longitudinal sampling vs. Illumina for trans-biome and circumboreal sampling), but we used these data to consider turnover (see Statistical analyses, EW transect, below).

QUANTIFICATION AND STATISTICAL ANALYSIS

Richness and phylogenetic diversity, trans-biome sampling

Read counts among SN 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 species (after pooling replicate microsites for each host species in each subsite) to the lowest number of sequences (i.e., 17,672) following recommendations by Weiss et al.[75] (see Figure S1). The *Cladonia* collection from SN5W yielded <5,000 reads from all microsites; thus, it was removed from subsequent analyses involving rarefied data. Calculation of OTU richness estimators and rarefaction analyses were done with the *vegan* package[40] in R[39] with rarefied data.

Multiple regression analyses of endophyte richness with *Picea* age and biomass, climate factors (PCA1, per above), nutrients in host tissues, and forest characteristics were carried out in JMP (versions 12 and 13. SAS Institute, Cary, NC, USA) (see Tables S2 and S5). We followed the principles of U'Ren et al.,[4] Zimmerman et al.,[76] and Barge et al.[77] in distinguishing spatial vs. environmental factors. In analyses presented in Table S2, we considered the relationship of endophyte richness to nutrients in host tissues and forest characteristics in analyses with and without accounting for an influence of climate. To consider the influence of climate in these cases, we used linear regression of endophyte richness (log transformed) with climate (PCA1) to generate residuals, which we then used as the dependent variable in regressions with nutrients from host tissues and forest characteristics (Table S2). The approach of analyzing residuals precludes including climate statistics in the usual way (e.g., with climate (PCA1) as an explanatory factor listed as a row on the table), but all relevant statistics are available in Table S2. In analyses presented in Table S5, we included climate (PCA1) as an explanatory variable, and thus PCA1 appears accordingly. We used the R package *phyloseq*[41] to extract the number of fungal classes represented by endophyte communities in each host as an estimate of phylogenetic diversity (short sequencing reads of the variable ITS2 region precluded robust phylogenetic analyses, as discussed in U'Ren et al.[4]).

Endophyte community structure

A Hellinger transformed distance matrix, constructed after removing rare OTU (i.e., OTU < 25 reads), was used as the input for the "betadisp" function in the *vegan* package[40] in R[39] to quantify the degree of multivariate beta dispersion among circumboreal vs. trans-biome endophyte communities from each of the three focal hosts. Venn diagrams and total Hellinger distance for circumboreal vs. trans-boreal endophyte communities were computed for each host genus using all OTU, including singletons. To quantify differences in beta diversity due to differences in climate at the southern and northern extremes of the transect, we calculated pairwise Hellinger distances between hosts from each site to hosts in SN1 (i.e., "southern"), SN9 (i.e., "northern"), and the closest site(s).

We visualized the effect of host on endophyte community structure along the SN gradient via NMDS ordinations based on Hellinger distance after removing OTU with < 25 reads. PERMANOVAs, implemented using the "adonis" function in *vegan*,[40] were used to assess the statistical significance of host species on endophyte community composition. Pairwise values of Hellinger distance between combinations of different host genera in each site (e.g., *Picea* vs. *Cladonia*) were calculated to assess the degree to which interspecific differences in endophyte communities shifted along the SN transect.

To identify the number of distinct endophyte communities present in each host species along the SN gradient, we used the Latent Dirichlet Allocation (LDA).[9] LDA is a probabilistic model that represents gradual changes in community composition while allowing for missing data and estimates of uncertainty.[9] We identified the most likely number of component communities for each host species

Please cite this article in press as: U'Ren et al., Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.01.063

based on the Akaike information criterion (AIC) for models with different numbers of component communities (chosen *a priori*).[9] LDA assigns each component community a value of theta, which indicates the proportion of each component.

Spatial autocorrelation and distance-based redundancy analyses (dbRDA)

We computed Mantel correlograms of Hellinger community distance and intersite geographic distances to quantify spatial autocorrelation for endophytes of each host.[78] Intersite distances were measured with the Haversine method in the R package *fields*.[42] Correlation coefficients were computed after 999 permutations. We plotted pairwise endophyte community Hellinger distances and intersite distances to assess potential distance decay. To test the significance of site and environmental variables 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[42,79–81] (Table S3). For the SN transect, analyses were conducted with and without detrending the Hellinger transformed OTU matrix by latitude, as recommended by Legendre and Legendre[82] when the response variable represents a linear gradient in one or two geographic dimensions (Table S3). However, detrending the community data spatially removed any significant effect of climate. The "ordiR2step" function in *vegan* was used for forward model choice solely on adjusted R^2 and P-values. We also used dbRDA to assess variation attributable to spatial eigenvectors alone (Table S3).

Endophyte networks

We assessed the degree of host specialization for endophyte communities across the SN gradient with bipartite networks calculated with the package *bipartite* in R[43,83] (Table S4). We calculated the frequency-based specialization index, H_2' for all hosts in each subsite using the "H2fun" function. H_2' is a measure of specialization generalized across the entire network.[83] Values of H_2' are based on the potential associations given the abundance of OTU and range from 0 (no specialization) to 1 (perfect specialization). At each site we observed a consistently high level of network specialization for each host genus (average value of H_2' for all hosts after Hellinger transformation: 0.83 ± 0.04; Table S4).

Using the "dfun" function, we also calculated the species-level specialization index d' for each host separately at each site (Table S4). The d' index calculates how strongly a species deviates from a random sampling of available interacting partners.[83] Values of d' range from 0 (generalized network) to 1 (specialized network). For each network, replicate samples for each host species were collapsed and the matrix was transformed with Hellinger prior to network calculations. Endophytes of all hosts displayed specialized networks and we observed no significant difference in d' values for each host species along the SN transect (Table S4).

Analyses of nestedness and turnover

For each host we used the R package *betapart*[44] to assess whether differences in endophyte communities across the SN gradient were related to nestedness or turnover, or combinations of both processes (Table S4). Nestedness occurs when OTU in sites with lower richness are subsets of the OTU found at sites with higher richness, whereas spatial turnover implies the replacement of some species by others as a consequence of environmental sorting or spatial and historical constraints.[84] The rarefied OTU matrix was transformed to presence/absence data and the Sorenson index was used with the function "beta.multi" to calculate the turnover component (measured as Simpson dissimilarity), the nestedness component (measured as nestedness-resultant fraction of Sorensen dissimilarity), and the overall beta diversity (measured as Sorensen dissimilarity).

Statistical analyses, EW transect

Read counts among EW 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 species (after pooling replicate microsites for each host species in each subsite) to the lowest number of sequences (i.e., 5,559) following recommendations by Weiss et al.[75] To visualize endophyte community structure across in endophytes of *Picea* and *Pleurozium* along the EW transect we used NMDS ordinations based on Hellinger dissimilarity after removing rare OTU (i.e., OTU < 25 reads) in conjunction with PERMANOVA to assess the statistical significance of host species in defining endophyte community composition (Table S3). PERMANOVA were implemented using the "adonis" function in *vegan*. Distance decay and Mantel correlograms were conducted as described above for the SN transect. We used dbRDA with PCNM to assess variation attributable to spatial eigenvectors alone (Table S3). We also tested the significance of sequencing run and environmental variables on endophyte communities while constraining variation attributable to distance or run alone. After accounting for sequencing run or significant eigenvectors (*Pleurozium* only), we observed no significant differences in community composition among endophytes of *Picea* or *Pleurozium* as a function of site or climate variables (MAT or MAP) across the EW transect (Table S3). Two spatial eigenvectors were significant for *Pleurozium* in PCNM (Table S3), which together explained 8% of the variation in endophyte community composition with dbRDA. However, variation partitioning analysis[85] revealed that sequencing run explained more variation than spatial eigenvectors (run adjusted R^2 = 0.08; PCNM10 adjusted R^2 = 0.04; residuals = 0.88). Variation partitioning was implemented with the function "varpart" in *vegan*. Network analyses, including host specialization, species-level specialization index d' for each host, nestedness and turnover were performed for endophyte communities along the EW transect as described above (Table S4).

Current Biology, Volume *34*

Supplemental Information

Environmental drivers and cryptic biodiversity hotspots define endophytes in Earth's largest terrestrial biome

Jana M. U'Ren, Shuzo Oita, François Lutzoni, Jolanta Miadlikowska, Bernard Ball, Ignazio Carbone, Georgiana May, Naupaka B. Zimmerman, Denis Valle, Valerie Trouet, and A. Elizabeth Arnold

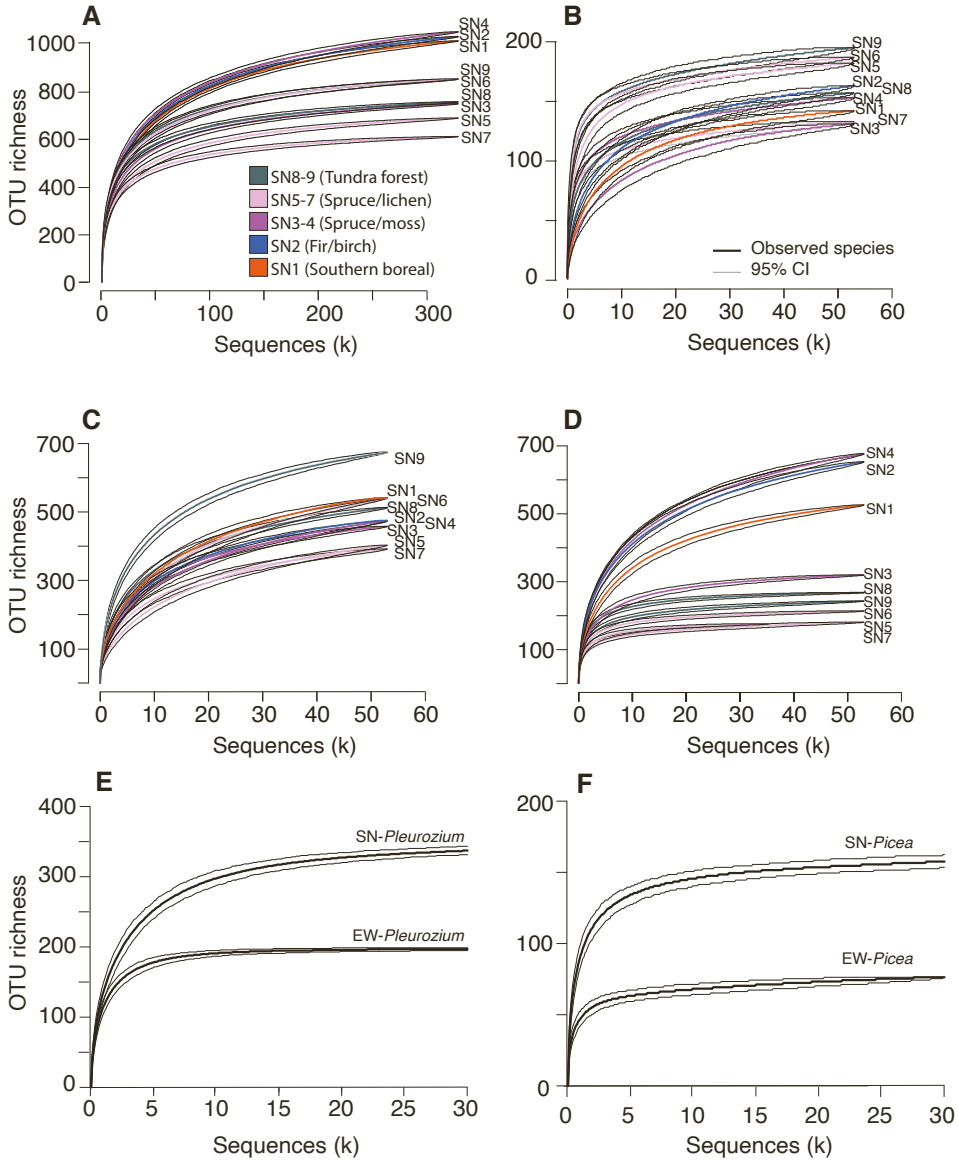

Figure S1. Rarefaction of endophytic fungal communities associated with boreal hosts, Related to STAR Methods. **(A)** All sites across the transboreal SN gradient, and each focal host shown separately on that gradient: **(B)** *Picea*, **(C)** *Pleurozium,* and **(D)** *Cladonia*; **(E)** comparison of transboreal (SN) and longitudinal (EW) gradient for *Pleurozium*, and **(F)** for *Picea*. For panels **A-D**, lines are colored according to Figure 1A, and each sample was subsampled to 17,672 reads (after removal of samples with < 5000 reads). For panels **E-F**, data from the SN transect were sub-setted to match the same number of samples as the EW transect, and OTU with < 25 reads were removed from the analyses (see Methods; results are similar when all OTU were included). Thin grey lines represent 95% confidence intervals around the means.

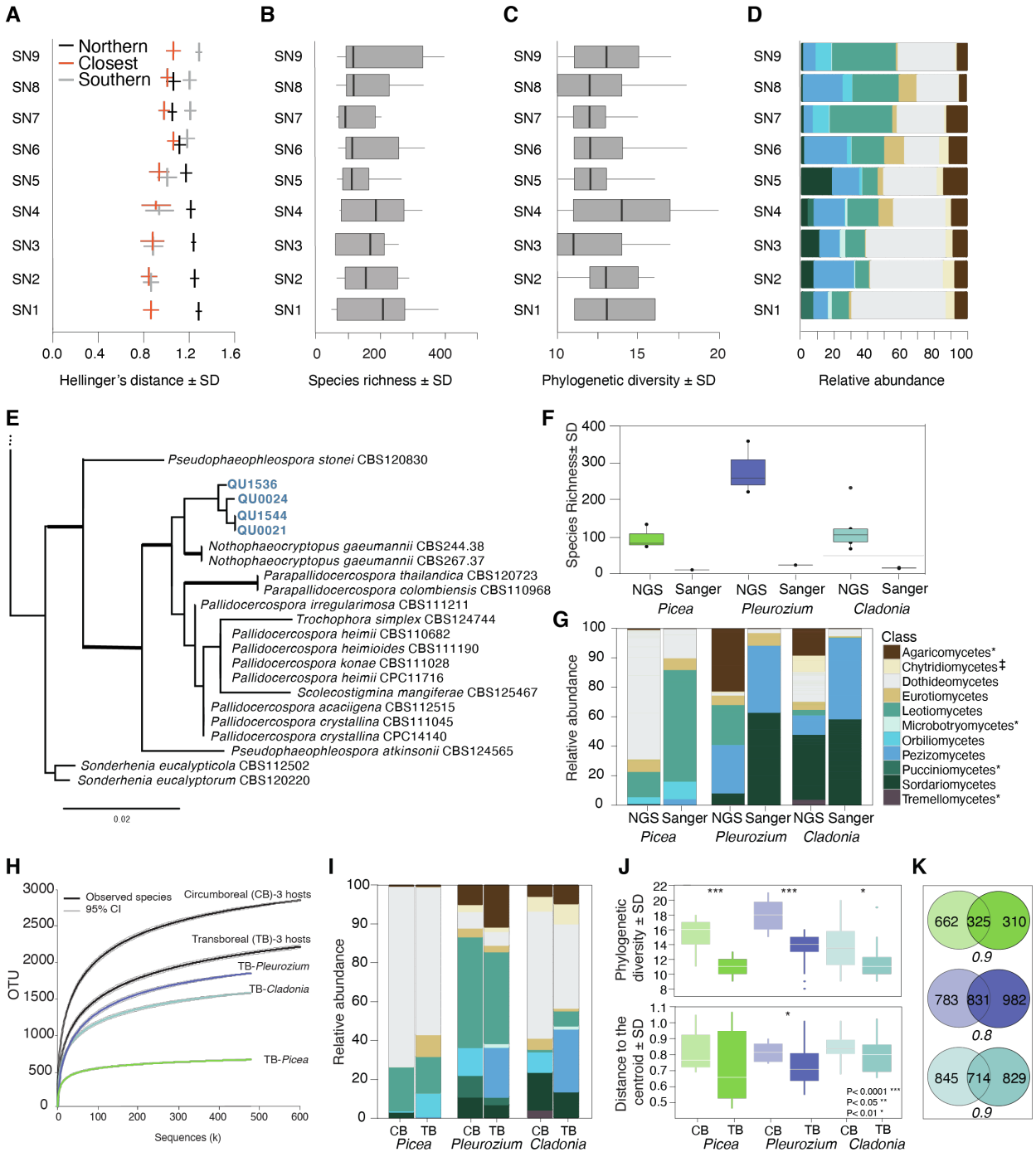

Figure S2. Summary of transboreal data for all host species from culture-free Illumina analyses and culture-based collections, and comparison of transboreal and circumboreal data sets, Related to Figure 2, Figure 3, and STAR Methods. (**A**) Beta diversity (mean ± SD), (**B**) species richness, (**C**) phylogenetic diversity, and (**D**) phylogenetic composition of endophytes from three hosts at a transboreal scale. Phylogenetic diversity was calculated as the number of fungal classes per sample. Colors for fungal classes follow Figure 2B. Beta diversity was calculated as Hellinger's distance between endophyte communities in the closest site vs. the southernmost (SN1) and northernmost (SN9) site. For example, the endophyte community in the southernmost site, SN1 was more similar to that in SN2 (lower Hellinger's

distance) than in the northernmost site (higher Hellinger's distance). **(E)** Phylogenetic analysis of four cultured representatives of the dominant Capnodiales OTU in *Picea* from the transboreal transect (shown in blue) illustrates a sister relationship to the pathogen of Douglas fir, *Nothophaeocryptopus gaeumannii*. The tree represents a subset of the phylogenetic analysis of LSU nrDNA (see Methods). Thickened branches indicate bootstrap support > 70%. **(F-G)** Comparison of culture-free NGS and culturing on the richness (**F**) and taxonomic composition (**G**) of endophytes of *Picea, Pleurozium,* and *Cladonia* at SN5. All sequences and OTU were included in analyses. Taxonomic comparisons were limited to the most abundant classes for visualization purposes. Symbols denote non-Ascomycota classes (i.e., Basidiomycota: asterisk; Chytridiomycota: double dagger). Comparison with cultures revealed seventeen-fold greater richness obtained by NGS compared to culturing. When Illumina data were subsampled to match the total number of sequenced cultures per sample, NGS provided an approximately twofold increase in richness relative to culturing. **(H)** Richness from transboreal sampling (TB) in Québec (SN1-SN9) vs. circumboreal sampling (CB) of the same host genera (*Picea, Pleurozium, Cladonia*) across Eurasia and North America, with the same methods for culture-free analyses[4]. CB methods and details are in ref. 4. **(I)** Composition of endophyte communities obtained by barcode amplicon sequencing for TB vs. CB (symbols denote non-Ascomycota). **(J)** Phylogenetic and beta dispersion (asterisks indicate significant differences, CB vs TB). **(K)** For CB (lighter) and TB (darker), distinctive and shared species and Hellinger community distance (below circle).