nature portfolio

Corresponding author(s): Jordan Peccia and Jonathan Gewirtzman

Last updated by author(s): May 29th, 2025

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Software was not used for data collection.
Data analysis	Pre-processing Nephele v2.23.2 104, using the DADA2 105 R package v1.28, was used for pre-processing of sequence data, including filtering, trimming, merging of paired-end sequences, chimera removal, and taxonomic assignment. For 165 sequencing, forward reads were truncated to 260 bp, reverse reads were truncated to 180 bp, maxEE was set to 5, truncQ was set to 4, and the SILVA 106 v138.1 database was used for taxonomic assignment. The median read count for 16S data following pre-processing was 10,014 reads, with 86.7% of ASVs resolved at the class level. For ITS data, primers were removed using cutadapt v4.5, maxEE was set to 5, truncQ was set to 4, min_len was set to 50, and taxonomic assignment was done using the UNITE v8.3 database. The median read count for ITS data following pre-processing was 21,957 reads, with 52.8% of ASVs resolved at the class level. Post-processing Following pre-processing, sequence data was imported into R v.4.3.0 for analysis using phyloseq 107 v.1.44.0 and microeco 108 v.0.17.0. An unrooted phylogenetic tree was used for ITS data, while a rooted tree was used for 16S data. For ITS sequence data, samples were rarefied to a depth of 4,000 reads (SI Figure 14). For 16S sequence data, remnant chloroplast and mitochondrial ASVs were first filtered out (10.3% of ASVs), and then samples were rarefied to a slightly lower depth (3,500 reads) to account for sequence loss during plastid filtering. Rarefaction curves were produced by Nephele using plot.ly v.2.5.1. For source-tracking analyses, FEAST 109 v.0.10 was used. All analyses, unless otherwise specified, were paired (at the level of the individual tree), with Expectation-Maximization iterations set to 10,000. Taxa were agglomerated at the family level for the whole-forest study, whereas they were agglomerated at the genus level for the black oak study. For construction of the phylogeny of tree species, the phytotools 110 package v.1.5-1 and S.PhyloMaker function were employed, drawing on the PhytoPhylo megaphylogeny as a backbone. For maps of tree genera distribution, the rnaturalearth 111 v.0.3.3 package was used for

illustration of data sourced from the GlobalTreeSearch database 112. For broad-level ordinations (e.g., between different sample types), dimensional reductions were performed using principal component analyses (PCoAs) on weighted unifrac distance. For within-sample-type ordinations, where differences between sample communities were expected to be finer, dimensional reduction was performed on unweighted unifrac distances using uniform manifold approximation and projection (UMAP) via the umap v.0.2.10 package 113. For differential abundance analyses, the LEfSe method was employed using microeco. For LEfSe analyses, the significance level was set to 0.01, and the threshold was set to 3.5. Taxa unresolved at the class level were removed. In the case of functional inference, the FAPROTAX v.1.2.6 database 114 was utilized for prokaryotic communities, whereas the FungalTraits 115 and FUNGuild 116 databases were used for fungal communities. For heatmaps based on functional inference, weighted abundances were used. Adonis2 PerMANOVA tests 117,118 were performed within microeco.

Code Availability
All code used in this analysis, along with all necessary files, are openly available on GitHub and can be accessed at the following URL: https://github.com/jgewirtzman/tree-microbiome.git.

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data Availability
The data that support the findings of this study have been deposited in the NCBI database under accession number PRJNA1124946. These data are publicly accessible at NCBI. For mapping global tree distributions, data from GlobalTreeSearch was utilized (https://www.bgci.org/resources/bgci-databases/globaltreesearch/) was used in manner pursuant to their non-commercial BGCI Data User Agreement.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

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

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

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

Ecological, evolutionary & environmental sciences study design

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

Study description	The study aimed to characterize the microbiome within the wood of living trees, focusing on the distinct microbial communities inhabiting heartwood and sapwood. The research was conducted using a hierarchical sampling design. Samples were collected from 150 trees across 16 species in the northeastern United States, representing a mix of deciduous and evergreen species. The primary treatment factors were the tree species and the type of wood tissue (heartwood vs. sapwood). The study included both qualitative and quantitative characterizations of microbial communities, employing novel sample processing methods and sequencing techniques to evaluate the diversity and abundance of prokaryotic and fungal taxa. Additionally, a single black oak tree was felled and sectioned into its component parts (individual tree tissues, soil layers, and

Jewwns 5upoda.a I oflop.Jod eJmeu
2
aboveground forest components) and the microbiomes of these components sequenced (also prokaryotic

Research sample	The research sample comprised 150 living trees from 16 species native to the northeastern United States, including Acer rubrum (red maple), Betula lenta (black birch), and Pinus strobus (Eastern white pine), among others. These species were chosen due to their wide distribution and ecological significance in the region. The trees varied in age and size, with a minimum diameter at breast height (DBH) of 51 cm, ensuring mature specimens. This sample represents a broad spectrum of tree species common in the Central Hardwood-Hemlock-Pine region, providing a comprehensive overview of the wood microbiome in this biome.
Sampling strategy	The sampling strategy involved collecting wood cores from each tree using increment borers. Four cores were taken from each tree at standardized heights to account for potential spatial variability within the tree. Soil samples were also collected from the rhizosphere of each tree to investigate potential microbial exchange between the soil and wood. Sample sizes were determined based on preliminary studies indicating significant microbial diversity and were deemed sufficient to capture the variability across different species and tissues. No formal sample-size calculation was performed.
Data collection	Data collection involved extracting wood cores and soil samples, which were then processed for DNA extraction and sequencing. Cores were collected during the peak growing season to ensure active microbial communities. Samples were immediately placed on dry ice and stored at -80°C until processing. DNA was extracted using a standardized protocol and sequenced for 16S rRNA and ITS regions to identify bacterial and fungal taxa. The data were recorded by trained researchers following rigorous standard operating procedures to ensure consistency and accuracy.
Timing and spatial scale	Data collection occurred between July 19 and August 12, 2021, for the whole-forest study, and on October 4, 2022, for the comprehensive sampling of a black oak tree. Samples were collected at different heights within the trees and from various soil depths to capture the vertical and horizontal diversity of the microbiome. The spatial scale encompassed the Yale-Myers Forest in northeastern Connecticut, with specific coordinates of 41.9° N, 72.r W.
Data exclusions	ITS samples were rarefied to 4,000 reads. 16S data were filtered for remnant chloroplast and mitochondrial ASVs and rarefied to slightly lower depth (3,500 reads) to account for sequence loss during plastid filtering. Samples below these read depths were excluded. No samples were otherwise or manually excluded from the dataset.
Reproducibility	To verify reproducibility, the study included multiple replicates for each tree species and tissue type. Sequencing and data analysis were conducted using established protocols with internal controls. All attempts to replicate the findings were successful, and consistent results were obtained across different sampling times and tree species.
Randomization	This is not relevant to the study; we surveyed the forest and sampled hierarchically, targeting a wide range of species. Species were sampled within species and DBH ranges by order of encounter, and the sampled cohort is distributed semi-randomly within the sampled stands.
Blinding	Blinding was not applicable during data acquisition due to the nature of field sampling.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	The field work was conducted during the peak growing season, from July 19 to August 12, 2021, and on October 4, 2022, for the comprehensive sampling of a black oak tree. The average temperature during the sampling period ranged from 20°C to 30°C, with occasional rainfall typical of the summer season in the northeastern United States.
Location	The study was conducted at the Yale-Myers Forest, a working and research forest located in northeastern Connecticut, USA (41.9529' N, 72.1239° W). This forest is classified in the Central Hardwood-Hemlock-Pine region and comprises mainly secondgrowth vegetation. The specific sampling locations within the forest were chosen to represent a wide range of tree species and environmental conditions typical of the region.
Access & import/export	All necessary permits and approvals were obtained to conduct the field work and collect samples at the Yale-Myers Forest. The research adhered to local, national, and international laws governing the collection and export of biological samples. Permissions were granted by the Yale School of the Environment, and no samples were imported or exported beyond the scope of the study. All sampling procedures were conducted responsibly to minimize impact on the environment and comply with legal requirements.
Disturbance	The study caused minimal disturbance to the forest ecosystem. Increment borers were used to collect wood cores, which allowed for precise and minimally invasive sampling. Soil samples were collected using a soil probe, ensuring that the sampling process did not significantly disrupt the soil structure or root systems. One single tree was felled for the study, in an area designated for future active management for maple syrup and timber production. The research team took care to leave the study sites as undisturbed as possible, ensuring that the forest ecosystem remained intact.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

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

Plants

Seed stocks	Non-living samples of plant tissues (wood, and additional aboveground tissues) were sampled following procedures described above.
Novel plant genotypes	N/A
Authentication	N/A