frontiers in Forests and Global Change

TYPE Original Research
PUBLISHED 08 May 2026
DOI 10.3389/ffgc.2026.1782331

OPEN ACCESS

EDITED BY
Ari Mikko Hietala,
Norwegian Institute of Bioeconomy Research (NIBIO), Norway

REVIEWED BY
Joseph Kawash,
Agricultural Research Service (USDA), United States
Andrew Pais,
Vance-Granville Community College, United States

*CORRESPONDENCE
Andrew Miles
✉ miles.375@osu.edu

†These authors have contributed equally to this work

RECEIVED 06 January 2026
REVISED 31 March 2026
ACCEPTED 02 April 2026
PUBLISHED 08 May 2026

CITATION
Miles A, Torres-Bedoya E, Conrad AO and Bonello P (2026) Foliar infrared spectra track nematode density and symptom-specific phytobiome signatures in beech leaf disease. *Front. For. Glob. Change* 9:1782331. doi: 10.3389/ffgc.2026.1782331

Foliar infrared spectra track nematode density and symptom-specific phytobiome signatures in beech leaf disease

Andrew Miles[1]*†, Eliana Torres-Bedoya[1†], Anna O. Conrad[2] and Pierluigi Bonello[1]

[1]Department of Plant Pathology, The Ohio State University, Columbus, OH, United States, [2]Northern Research Station, USDA Forest Service, Delaware, OH, United States

Beech leaf disease (BLD), caused by the nematode *Litylenchus crenatae* ssp. *mccannii* (LC), poses a severe threat to American beech (*Fagus grandifolia*) across eastern North America. The disease causes wholesale anatomical, morphological, and physiological alterations, including symptoms such as leaf banding and bud abortion; the latter eventually leads to beech mortality, especially of younger trees. This outcome severely affects regeneration and diminishes the important ecosystem services provided by this keystone species. To advance our understanding of the disease, we hypothesized that significant links exist among LC abundance, foliar phytobiome dysbiosis, and phytochemical alterations detectable via NIR reflectance spectroscopy. To test this hypothesis, we applied molecular diagnostics to quantify LC in the tissues, bacterial and fungal foliar microbial community profiling, and NIR reflectance spectroscopy to characterize three distinct tissue phenotypes in BLD-infected trees: (1) asymptomatic tissues of asymptomatic leaves (AA); (2) asymptomatic tissues of symptomatic leaves (AS); and (3) symptomatic (galled) tissues of symptomatic leaves (GS). Overall, the three tissue types differed significantly in LC load (AA < AS < GS). Furthermore, NIR spectral profiles differed consistently among tissue types, with distinct wavelength regions associated with water and structural chemistry driving the separation. Machine learning and multivariate models of NIR spectra predicted both LC abundance and bacterial community composition by tissue type, enabling possible discrimination of dysbiotic foliar phytobiomes, with moderately high accuracy, but not so for fungal community composition. Bacterial taxa such as *Pseudomonas*, *Wolbachia*, *Luteibacter*, and *Pedobacter* were significantly associated with LC infection. Taken together, these results validate our hypothesis. This study establishes NIR technology as a platform for LC quantification and, more broadly, as a tool for assessing bacterial dysbiosis in plant systems.

KEYWORDS
American beech, beech leaf disease, biosis, *Litylenchus crenatae* ssp. *mccannii*, near-infrared, pathogen detection, phytobiome

Introduction

Forest tree diseases caused by invasive pathogens are reshaping ecosystem structure and function worldwide, often outpacing both early detection efforts and mechanistic understanding of disease progression (Bonello et al., 2020; Guo et al., 2023; Paap et al., 2020; Williams et al., 2023). One prominent current example is beech leaf disease (BLD) (Ewing et al., 2019), an emerging epidemic affecting American beech (*Fagus grandifolia*) from the Great Lakes to the northeastern United States and Ontario, Canada (Kantor et al., 2025; Reed et al., 2020). BLD is caused by the foliar, migratory nematode *Litylenchus crenatae* ssp. *mccannii* (LCM, Anguinidae). It is a highly concerning disease because American beech, a late-successional hardwood tree, plays a foundational role in the local mixed hardwood forest by contributing to forest structure, nutrient dynamics, and the production of hard mast that supports diverse wildlife communities (Ross, 2013; Stephanson and Coe, 2017).

The epidemic nature of BLD is likely due to the fact that LCM is a non-native pathogen, probably originating in Japan, where it is associated with Japanese beech, *Fagus crenata* (Burke et al., 2020; Carta et al., 2020; Kanzaki et al., 2019; Selvi et al., 2025; Vieira et al., 2023). The emergence of BLD in North America, therefore, represents a jump to a new host species, from coevolved Asian hosts to non-coevolved American beech (Denk and Grimm, 2009), by a nematode that appears to be specialized to the genus *Fagus*. For this reason, the *mccannii* ssp. designation has been called into questioned (Paulo Vieira, personal communication). Accordingly, here we will refer to *Litylenchus crenatae* simply as LC.

American beech is environmentally constrained due to its sensitivity to water balance and temperature, and especially winter precipitation (Selvi et al., 2025), but its adaptability to low-light conditions in the understory has contributed to its ecological prominence, particularly as other major timber species have declined (Albright et al., 2023; Collin et al., 2017). However, American beech is expected to be under severe pressure in the future, not only from BLD, but also from projected future climate conditions (Selvi et al., 2025), in addition to a well-established but slow-moving fungal epidemic, beech bark disease (Cale et al., 2017). The rapid spread of BLD raises concerns about long-term impacts on forest composition and stability, as well as the potential risk to related beech species, e.g., in Europe (Balla et al., 2021; Colbert-Pitts et al., 2025; Kantor et al., 2025; Pramreiter and Grabner, 2023).

BLD's characteristic symptom is interveinal dark greening (banding) of leaves (Figure 1), a phenotype associated with elevated chloroplast and chlorophyll content and altered leaf development and physiology (Ewing et al., 2019; McIntire, 2023; Miles, 2023; Vieira et al., 2023). Recent studies have shown that LC infects developing buds, disrupting cellular architecture and perturbing cell cycle regulation in leaf primordia, leading to developmental and morphological abnormalities that result in the observed banding and other symptoms, including leaf 'crinkling', post-budbreak (Fearer et al., 2022b; Vieira et al., 2023). In addition to dark greening, the 'bands' (henceforth, bands) are characterized by swelling due to abnormal cell proliferation and thickened cell walls, a form of galling (Arriola and Isaias, 2021; De Freitas De Freitas Silva et al., 2025; Kanzaki et al., 2019; Sánchez-Moreno and Ferris, 2025; Viana et al., 2013). As might be expected, nematodes are also most abundant in these tissues (Vieira et al., 2023).

The intensity of symptoms and the elevated pathogen load observed in galled areas must result in differentially altered biochemical and phytobiome profiles in banded and unbanded areas of symptomatic leaves relative to tissue on asymptomatic leaves; the alterations would also likely be directly proportional to

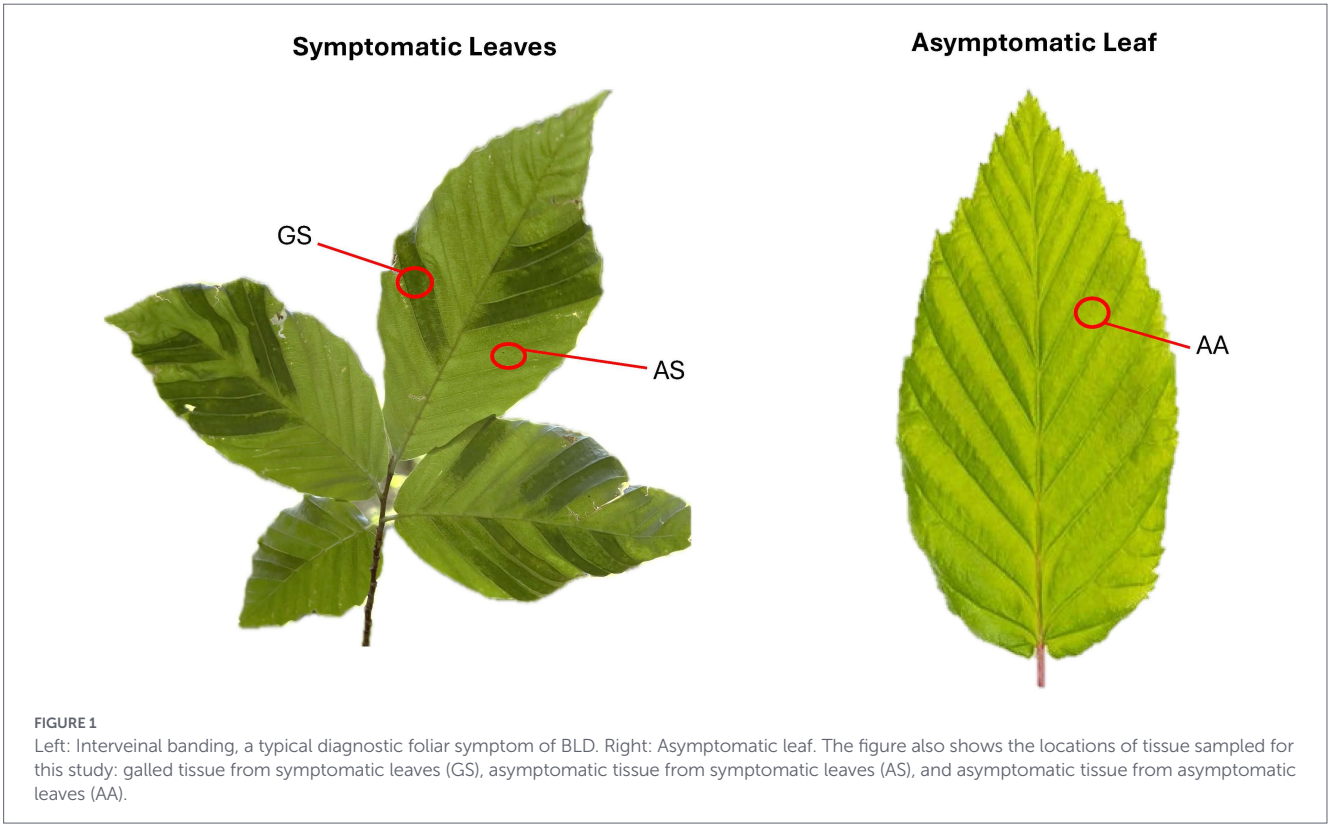

FIGURE 1
Left: Interveinal banding, a typical diagnostic foliar symptom of BLD. Right: Asymptomatic leaf. The figure also shows the locations of tissue sampled for this study: galled tissue from symptomatic leaves (GS), asymptomatic tissue from symptomatic leaves (AS), and asymptomatic tissue from asymptomatic leaves (AA).

nematode loads. Foliar phytobiomes associated with BLD have been studied before at a coarse, whole-leaf level (Burke et al., 2020, 2024; Ewing et al., 2019), but no integrated studies are available on how nematode load affects foliar chemistry and phytobiomes, and whether these changes can be detected rapidly and non-destructively, e.g., by using near-infrared (NIR) spectroscopy.

Near-infrared (NIR) spectroscopy measures reflectance or absorbance across hundreds of wavelengths, capturing biochemical fingerprints arising from the diverse vibrational states of organic molecule bonds and revealing mixtures of functional groups or metabolite classes. These signatures enable inference of metabolic changes linked to biotic or abiotic stress (Conrad and Bonello, 2016; Masaitis et al., 2013; Yu et al., 2022). Combined with advanced statistical methods or machine learning (ML) like random forest, support vector machine, or partial least squares, NIR becomes a powerful tool for characterizing tissue chemistry and classifying biological states (Boateng et al., 2020; Esposito Vinzi and Russolillo, 2013; Hardoim et al., 2016; Lu and Rasco, 2012; Türker-Kaya and Huck, 2017). The approach is becoming increasingly user-friendly, with handheld spectrometers that enable rapid, cost-effective phenotyping of disease resistance and pathogen load, including in the field. In the BLD context, NIR reflectance has been used to detect infected but asymptomatic beech leaves (Fearer et al., 2022a).

Host metabolism and chemistry, and microbial colonization, are tightly linked because microbial communities influence host biochemistry through both microbe-elicited host responses and direct microbial synthesis of a variety of metabolites (Pang et al., 2021). Thus, changes in spectral signatures reflecting biochemical alterations may also reveal shifts in the associated phytobiome (Hassani et al., 2018; Korenblum and Aharoni, 2019; Zhalnina et al., 2018). For example, studies in humans have shown that dysbiosis (abnormal or unhealthy microbiomes) can produce detectable spectral changes that precede or extend beyond dysbiosis-associated symptoms, indicating that spectral signatures can encode information about microbial community states (Ho et al., 2019; Li et al., 2025; Liu et al., 2022). However, the extent to which such relationships apply to plant systems remains largely unexplored.

Prior BLD phytobiome studies using 16S and ITS metabarcoding have identified distinct bacterial and fungal communities associated with symptomatic leaves. Burke et al. (2020) reported that bacterial, but not fungal, communities, differed significantly between symptomatic and asymptomatic leaves, with *Wolbachia* and *Mucilaginibacter* enriched in symptomatic leaves. Ewing et al. (2021) expanded this understanding, identifying the bacterial genera *Wolbachia*, *Erwinia*, *Paenibacillus*, and *Pseudomonas,* together with the fungal genus *Paraphaeosphaeria,* among the indicator taxa associated with symptomatic leaves. This work revealed that bacterial communities varied with symptom type, while fungal communities were shaped by both geography and symptom expression. More recently, Burke et al. (2024) demonstrated that nematode abundance correlated with fungal, but not bacterial, community shifts, highlighting potential cross-kingdom interactions within the diseased leaf environment. Despite these advances, studies of the BLD phytobiome have not explicitly integrated host biochemical phenotypes or non-destructive diagnostic approaches. It remains unclear whether host spectral traits can serve as proxies for microbial community assembly or pathogen titer in BLD, especially at the fine spatial scale achievable with handheld NIR spectroscopic tools.

Building on this foundation, we investigated how nematode abundance (load), foliar phytobiome organization, and host spectral traits are linked in BLD. We focused on three foliar symptom types: asymptomatic areas of asymptomatic leaves (AA), asymptomatic areas of symptomatic leaves (AS), and galled areas of symptomatic leaves (GS) (Figure 1). Integrating quantitative molecular diagnostics, bacterial and fungal community profiling, and NIR spectroscopy with multivariate statistics and ML approaches, we tested whether host spectral signatures can serve as non-destructive proxies for both pathogen burden and bacterial community assemblage. This framework moves beyond symptom-based diagnosis to provide new insights into more nuanced early detection, host–phytobiome–nematode interactions during BLD infection, and to evaluate the potential of NIR spectroscopy to infer microbial community dynamics and dysbiosis in plants.

Materials and methods

Study sites, plant material, and spectral data collection

This study was initiated in September 2025, utilizing a preexisting Allegheny National Forest (ANF) plot network established in 2023 as part of ongoing research on BLD (Miles, 2023). The network comprises 48 sites (Supplementary Table S1; Figure 2) established for long-term ecological monitoring associated with the Ash Network and the Kinzua Quality Deer Cooperative plot networks. For this study, three sites (S11, S14, and S16) were selected for tissue collection (Figure 2). Sites were chosen to maximize spatial separation and thus capture geographic variability (Supplementary Table S2; Supplementary Figure S1). Disease severity at each site was assessed according to the protocol outlined by Miles (2023), using a composite 0–16 index derived from the percentages of banded and crinkled leaves and the canopy vigor score for a single tree. Because (1) molecular validation of LC presence and load occurred only after tissue collection and spectral acquisition, (2) localized or systemic host responses may occur even at low or undetectable LC loads, and (3) BLD is highly prevalent in the study area, even AA tissues could not be assumed to be truly naïve for this study. However, including truly naïve trees as controls from outside the infestation zone (e.g., Central Ohio), likely would have introduced spurious variation in phytobiome composition driven by location (Ewing et al., 2021). Thus, our study did not include naïve trees.

Beech trees were randomly selected within a 50-meter radius of each site center. From each tree, five symptomatic and five completely asymptomatic leaves were collected for both spectral and tissue analysis. Leaves were haphazardly sampled from the lower to mid-canopy and were visually inspected to ensure the absence of physical damage, debris, lesions, or chlorosis and necrosis unrelated to typical BLD symptoms.

NIR reflectance spectra were collected from each symptom type (AA, AS, GS, Figure 1) using a NeoSpectra handheld spectrometer (Si-Ware Systems, La Cañada, CA, USA) with a coverage area of ~78.54 mm^2. Spectra comprised 257 bands spanning 1,350–2,550 nm and were acquired from the adaxial surface of leaves while avoiding the veins. A single scan was collected per sampling point, with a 2-s scan time and a 5-s interval between consecutive measurements. Spectral data were transferred via Bluetooth to the NeoSpectra

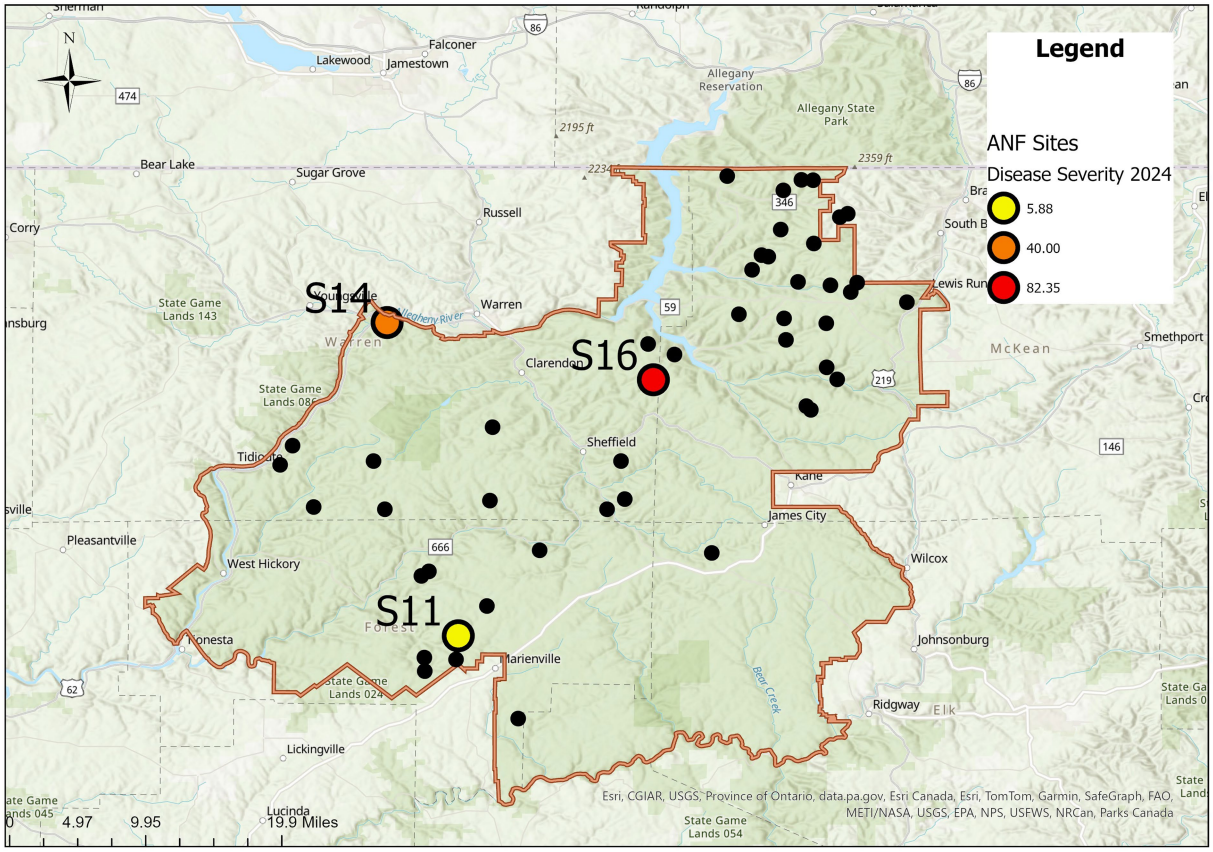

FIGURE 2
Locations of field sites (S11, S14, S16) sampled for this study within the Allegheny National Forest (ANF), Pennsylvania, USA. Points are colored in relation to the mean BLD disease severity recorded in 2024 and calculated as described in Miles (2023), with disease severity increasing from yellow to red. The black points represent the other sites in the plot network, with the ANF outline in orange. Map generated using ArcGIS Pro (Esri, Redlands, CA, USA). All site GPS coordinates are shown in Supplementary Table S1.

"Collect" mobile application, with blank reference scans performed at regular intervals for calibration.

Following NIR data collection, leaf disks were excised from interveinal regions using a 1/4-inch MS Standard Tissue Punch (MIDCO Global Inc., Kirkwood, MO, USA). Sampling was conducted across 10 total leaves per tree: five symptomatic and five asymptomatic. From the five symptomatic leaves, four disks were haphazardly sampled per interveinal area for both the GS and AS symptom categories. From the five asymptomatic leaves, four disks were sampled for the AA category. This process yielded a total of 20 representative leaf disks for each of the three tissue types per tree (Figure 1). The leaf punch was cleaned with 70% ethanol between excisions. Tissues were pooled by tree and symptom type, yielding 90 composite samples in total (3 sites × 10 trees × 3 symptom types). Samples were immediately flash-frozen in liquid nitrogen in the field and thereafter stored at −80 °C until DNA extraction.

DNA extraction, library preparation, and sequencing

Total DNA was extracted from leaf material using a modified two-day CTAB protocol based on Torres-Bedoya et al. (in preparation). Briefly, leaf tissues were ground to a fine powder in liquid nitrogen with Celite® 545 (Sigma-Aldrich) as a carrier. Approximately 25 mg of frozen tissue per sample was combined with ten 2.4-mm metal beads, 980 µL of CTAB extraction buffer (44 mM Tris, 56 mM Tris–HCl, 1.4 M NaCl, 50 mM Na_2EDTA-H_2O, 2.5% w/v CTAB, 1% w/v PVP-40), and 20 µL of 2-mercaptoethanol. Samples were vortexed for 1 min, incubated at 65 °C for 30 min with agitation every 10 min, then mixed with 500 µL chloroform:isoamyl alcohol (24:1) and centrifuged at 14,000 rpm for 5 min. The aqueous phase was transferred to a new tube, precipitated with an equal volume of cold isopropanol, and incubated overnight at −20 °C. The following day, DNA was pelleted by centrifugation (14,000 rpm, 1 min), washed with 1 mL of 70% ethanol, air-dried, and resuspended in 50 µL of TE buffer. DNA concentration and purity were assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, U. S. A.). Amplifiability was verified by PCR of the full ITS region (ITS1-ITS4), and amplicons were visualized by agarose gel electrophoresis.

Equimolar amounts of DNA were used for library preparation and sequencing by the Molecular and Cellular Imaging Center (MCIC) at The Ohio State University. Libraries targeted the fungal ITS1 region using barcoded primers ITS1F and ITS2R, and the V4 hypervariable region of the 16S rRNA gene using primers 515F and 806R. Each PCR reaction contained 5 ng of template DNA, 0.2 µM of each primer, 1 × Phusion HF reaction buffer (New England Biolabs), 0.2 mM dNTPs (New England Biolabs), and 0.2 µL of Phusion HF DNA polymerase (New England Biolabs). Amplification was performed with an initial denaturation at 96 °C for 3 min; 30 cycles of 96 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min; followed by a final extension at

72 °C for 10 min. Amplicons were purified using AMPure XP beads (Beckman Coulter, USA). Indexing PCRs were conducted with the same reaction components and concentrations, using the following thermal profile: 96 °C for 3 min; 8 cycles of 96 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s; and a final extension at 72 °C for 10 min. Indexed amplicons were purified with AMPure XP beads, quantified, and pooled in equimolar concentrations.

Final library concentrations were determined using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, USA), and fragment size distribution was verified on a 4,200 TapeStation system (Agilent Technologies, USA). Sequencing was performed on an Illumina NextSeq 1000/2000 platform (Illumina Inc., San Diego, CA, USA) to generate 2 × 300 bp paired-end reads. The raw data were deposited in the NCBI Sequence Read Archive under BioProject accession numbers PRJNA1356492 (ITS) and PRJNA1356491 (16S).

LC quantification and data analysis

LC abundance was quantified using an in-house probe-based qPCR assay (Torres-Bedoya et al., in preparation). Briefly, to build a standard curve of nematode loads, whole nematodes in various numbers (0, 1, 10, 100, 1,000, and 2,000) were added to 25 mg FW ground leaf tissue collected from naïve American beech trees growing at the Chadwick Arboretum in Columbus, OH. As LC cannot be cultured *in vitro* at present, nematodes were extracted from field-collected bud and leaf tissues and enumerated prior to use. Standard curves were generated by spiking the known quantities of counted nematodes into naïve, LC-free beech tissue. Nematode densities used for calibration were determined via direct counts or volumetric subsampling of suspensions, averaged across replicates for counts exceeding 100 LC.

DNA was then extracted from the nematode-amended samples and diluted 1:15 v/v in molecular-grade water. The FAR-based qPCR assay utilized the forward primer FAR-F (5′-GAA GCC CAA AGG AGA TGA GAA G-3′) and reverse primer FAR-R (5′-TCC TCG GAC AGA GCC TTA TAC-3′) described by Vieira et al. (2023), along with a hydrolysis FAR-Probe (5′-FAM-AAA CCT GAA CAA GCT CCG CGA AGA-BHQ-1-3′). Quantification of standards and samples was based on triplicate qPCR reactions with non-template water controls included in each run. Only samples with consistent amplification across technical replicates were retained for analysis.

Differences in LC abundance among symptom types were assessed using one-way analysis of variance (ANOVA), followed by pairwise comparisons with Tukey's Honestly Significant Difference (HSD) test. Effect size was estimated using eta-squared (η^2), calculated from the ANOVA sums of squares. Similarly, differences in disease severity among sites were analyzed using ANOVA to test the effects of site, symptom type, and their interaction. Post-hoc pairwise comparisons were performed using Tukey's HSD test. The relationship between LC and disease severity at the site level was evaluated using linear models.

Spectral data processing and multivariate analysis

Spectral data were processed in R (v4.3.1). Raw reflectance curves were first visualized to assess baseline quality. Outlier detection was performed using the functional depth approach implemented in the *fda.usc* package (Febrero-Bande and de la Fuente, 2012), applying a 6% trimming level as described by Conrad et al. (2020). Spectra flagged as potential outliers were visualized and evaluated for anomalies. All spectra were then subjected to scatter correction using a standard normal variate (SNV) transformation implemented in the *prospectr* package (Stevens and Ramirez-Lopez, 2025). Corrected spectra and the subset of flagged curves were re-plotted for comparison.

Second-derivative transformation was performed using a Savitzky–Golay filter (polynomial order = 2, derivative order = 2, and window size = 11 points). For downstream analyses, second-derivative spectra were averaged by tree within each symptom type to obtain a representative tree-level spectrum.

Principal component analysis (PCA) was performed on centered and scaled tree-level second-derivative spectra to visualize dominant axes of spectral variation. To statistically evaluate differences in spectral signatures, Euclidean distance matrices were computed from the scaled spectra, and permutational multivariate analysis of variance (PERMANOVA; 999 permutations, *adonis*) was used to test effects of symptom type and site.

Sparse partial least squares discriminant analysis (sPLS-DA) was applied to classify symptom types based on NIR spectral data using the *mixOmics* package (Rohart et al., 2017). Pairwise analysis was conducted for AA and GS tissues, while global analysis included all symptom types. The predictor matrix consisted of 246 spectral wavelengths. Model tuning was performed using repeated M-fold cross-validation (5 folds, 100 repeats) to select the optimal number of components and the number of wavelengths retained per component. Final models were fit using optimal parameters (Bonello et al., 2025; Fearer et al., 2022a). Variable importance in projection (VIP) scores were calculated to identify wavelengths contributing most strongly to symptom type discrimination.

Amplicon sequence processing and phytobiome analyses

Raw paired-end FASTQ files were processed using the nf-core/ampliseq pipeline (v2.15.0) (Straub et al., 2020). Sequence quality was assessed using FastQC, and adapter trimming was performed with Cutadapt (Andrews, 2010; Martin, 2011), with a minimum primer overlap of 3 bp and a maximum error rate of 0.1. High-quality reads were denoised, merged, filtered, and classified with DADA2 to infer non-chimeric amplicon sequence variants (ASVs) (Callahan et al., 2016). Quality filtering was performed using a maximum expected error threshold of 2 per read and a minimum read length of 50 bp. For 16S data, reads were truncated at 210 bp (forward) and 190 bp (reverse). For ITS data, paired-end processing was conducted using the –illumina_pe_its option and truncation positions were automatically selected using a median quality threshold of 25 while retaining at least 75% of the reads. Taxonomic annotation of fungal ASVs was performed against the UNITE reference database (version 10) (Nilsson et al., 2019) and bacterial ASVs were classified using the SILVA database (version 138) (Quast et al., 2012).

Phytobiome community analyses were conducted in R using the *MicrobiotaProcess* (Xu et al., 2023) and *vegan* packages (Oksanen et al., 2025). Taxonomic tables were curated to remove Eukaryota, Archaea, and unknown kingdoms, as well as mitochondrial reads and poorly resolved or spurious bacterial and fungal phyla. To account for differences in sequencing depth, read counts were first rarefied for the calculation of α-diversity metrics, and rarefaction curves were generated. α-diversity was quantified using observed ASVs, Chao1, ACE, Shannon, Simpson, and Pielou's evenness indices.

In this study, dysbiosis was measured as a shift in microbial community composition driven by pathogen pressure that alters α-diversity patterns, rather than by the presence or absence of individual taxa. Consistent with the Anna Karenina Principle (Zaneveld et al., 2017), healthy phytobiomes are expected to exhibit constrained compositional similarity, whereas pathogen infection can lead to increased community dispersion and stochasticity (Arnault et al., 2023). Such changes in microbial community organization may also be tightly linked to the host's physiological state and, therefore, detectable and identifiable through spectral traits.

For α-diversity analyses, ASV relative abundances were Hellinger-transformed, and Bray–Curtis dissimilarities were computed from the transformed data. Differences in community structure were evaluated using PERMANOVA (adonis function, 999 permutations), with models including symptom type, site, and their interaction as explanatory variables. Principal Coordinates Analysis (PCoA) was used to visualize multivariate patterns in community composition and to illustrate the proportion of variance in Bray–Curtis dissimilarities attributable to these factors. Differential abundance analysis was conducted at the genus level. Statistically significant differences between groups were first identified using a non-parametric Kruskal-Wallis rank-sum test ($p < 0.05$). To quantify the magnitude of these differences, linear discriminant analysis (LDA) was used to calculate the effect size for each differentially abundant genus.

Indicator species analysis was performed at the genus level using the *multipatt* function in the *indicspecies* package (De Cáceres and Legendre, 2009). Associations were evaluated with the point–biserial correlation coefficient (r.g), with significance assessed by 999 permutations. Genera with permutation-based p-values < 0.05 and indicator values (IndVal) ≥ 0.450 were considered significant and strong indicators of individual symptom types or combinations thereof.

Phytobiome–LC abundance association analysis

Associations between microbial community structure and LC abundance were assessed for bacterial (16S) and fungal (ITS) datasets. Analyses were performed using Hellinger-transformed ASV abundance matrices paired with corresponding sample metadata. LC $\log_{10}$ (abundance+1)–transformed abundance was treated as a continuous predictor. Distance-based redundancy analysis (dbRDA) was conducted using Bray–Curtis dissimilarities (vegan:capscale) to quantify the proportion of compositional variation attributable to LC abundance. Model significance was evaluated using permutation tests (999 permutations), and the variance accounted for by the first constrained axis (dbRDA1) was derived from the constrained eigenvalues.

To evaluate the effect of LC abundance independent of symptom type, a partial dbRDA was performed with symptom type included as a conditioning variable. Significance of the LC term after conditioning was assessed using 999-permutation tests.

To identify microbial taxa associated with LC abundance, with a focus on bacterial communities, multivariable linear models were implemented using the *MaAsLin2* package (Mallick et al., 2021). Taxon abundance tables were normalized using total-sum scaling and log-transformed prior to analysis. Models included LC abundance as a fixed effect, with symptom type and site included as covariates. Multiple testing correction was performed with the Benjamini-Hochberg false discovery rate (FDR), and taxa with q-values < 0.05 were deemed significant. Significant associations were summarized by effect size and direction.

NIR–LC association analysis

Relationships between NIR spectra and LC abundance were evaluated using ordination, correlation, and regression approaches. Nematode counts, log-transformed as above, were used for all analyses. PCA was performed on scaled spectral data to visualize major changes in spectral variation in relation to nematode abundance and symptom type. Wavelength-specific associations were evaluated using Spearman's rank correlation coefficients between individual spectral bands and nematode abundance, with FDR correction. Predictive relationships with spectral data and LC abundance were further assessed using a partial least squares regression (PLS-R) implemented in the *pls* package (Liland et al., 2025). The optimal number of components was selected based on cross-validated root mean squared error of prediction (RMSEP), and model performance was assessed using the coefficient of determination (R^2), root mean square error (RMSE), and residual plots to assess homoscedasticity, bias, and normality. VIP scores were calculated to identify influential wavelengths.

Phytobiome–NIR association analysis

NIR spectra were used to predict bacterial community composition derived from 16S rRNA gene sequencing data. ASV tables were Hellinger-transformed and summarized using Bray-Curtis dissimilarities, followed by principal coordinates analysis (PCoA). Initial analyses focused primarily on PCo1, which captured the largest proportion of variation in community composition. Spectral data were partitioned into training and test sets (70 and 30%, respectively) using a fixed random seed. A standard 70/30 training/testing sample split was used to preserve an independent evaluation set while retaining an adequate training sample size.

Predictive models included Random Forest (RF) regression using the package *randomForest* (Breiman, 2002) using $n = 500$ trees (Liaw and Wiener, 2002), partial least squares regression (PLS-R) using the *MixOmics* package (Rohart et al., 2017), elastic net regression using the *glmnet* package (Friedman et al., 2010), and support vector machine (SVM) with radial kernels using the *e1071* package (Meyer et al., 2019). The radial kernel allows SVM to capture nonlinear relationships, which are common in spectral and ecological datasets (Du et al., 2024; Razaque et al., 2021). Model performance was evaluated using test-set R^2, RMSE, and average calibration error (ACE). VIP scores were extracted, and influential wavelengths were identified.

Given the superior performance of RF and SVM models, these approaches were used to jointly predict the first two PCoA axes, with performance assessed using joint R^2 and Euclidean error in ordination space. Observed and predicted phytobiome configurations were visualized in PCoA space to assess predictive accuracy.

Results

NIR spectral profiles of beech leaves

A total of 449 NIR spectra were obtained across all samples and symptom types. Reflectance profiles exhibited consistent structural patterns (Figure 3), with pronounced absorption features in the

1,400–1,500 nm and 1,900–2,000 nm regions, wavelengths typically associated with O-H stretching overtones and leaf water content (Huang et al., 2025). AA tissues exhibited higher reflectance and shallower absorption features across the 1,400–2,500 nm range, whereas GS tissues showed reduced reflectance and deeper absorption troughs. AS tissues displayed intermediate spectral profiles, partially overlapping with both AA and GS (Figure 3).

Twenty-six spectra were flagged as potential outliers. However, comparison of raw, SNV-corrected, and second-derivative profiles indicated that these deviations reflected biological variation among symptom types rather than technical artifacts (Supplementary Figure S2). Consequently, no spectra were excluded during quality control. Application of the second-derivative transformation produced the expected edge effects, resulting in the removal of five spectral bands at each end of the spectrum.

Pairwise PERMANOVA revealed significant differences in NIR spectral profiles among sites (all pairwise comparisons, $p = 0.003$). Identical p-values reflect the fixed permutation scheme used in the analysis (999 permutations), which limited the minimum attainable p-value. Additionally, the effect sizes were weak, with low R^2 values across contrasts: S14 vs. S16 ($F_{1,298} = 13.24$, $R^2 = 0.043$), S11 vs. S14 ($F_{1,299} = 13.80$, $R^2 = 0.044$), and S11 vs. S16 ($F_{1,298} = 9.71$, $R^2 = 0.032$), indicating limited spectral differentiation attributable to geographic location. In contrast, symptom type explained substantially greater spectral variation, with all pairwise comparisons having significant ($p = 0.003$) and larger effect sizes, with the strongest separation occurring between AA and GS tissues ($F_{1,298} = 173.10$, $R^2 = 0.368$), followed by GS vs. $F_{1,298} = 78.45$, $R^2 = 0.209$) and finally AA vs. ($F_{1,299} = 23.12$, $R^2 = 0.072$). These results demonstrate that symptom type is the dominant driver of spectral variation, with GS tissue exhibiting the most distinct spectral signature. Consistent with these findings, PCA ordinations (Figure 4) showed pronounced clustering by symptom type with samples forming a gradient from AA to AS to GS.

sPLS-DA identified discrete wavelength regions driving the strong spectral separation between AA and GS tissues. The final model retained 30 variables across two components, including 25 features for PC1 and 5 for PC2. The minimum average balanced error rate (BER) during tuning was 0.0135 for PC1, indicating high discriminatory performance. Permutation testing confirmed that classification accuracy exceeded that expected by chance. VIP scores revealed that wavelengths contributing most strongly to AA/GS separation clustered into five discrete wavelength regions, approximately 1,410–1,460, 1,680–1,730, 1,760–1,820, 1,920–1,980, and 2,270–2,340 (VIP > 1; Supplementary Figure S3). These regions correspond to the absorption features associated with leaf water status, structural carbohydrates, and mixed C–H and N–H bonds (Türker-Kaya and Huck, 2017; Weyer and Lo, 2002; Workman and Weyer, 2007), the latter probably related to differential nitrogen allocation among tissue types.

Global sPLS-DA models incorporating all symptom types also showed strong separation between AA and GS, with AS tissues occupying an intermediate position, consistent with a progressive biochemical transition across symptom types. Unlike the pairwise comparison of AA/GS, the global VIP profile exhibited a broader distribution of moderate-importance peaks, reflecting the inclusion of additional wavelengths required to discriminate AS tissues. The

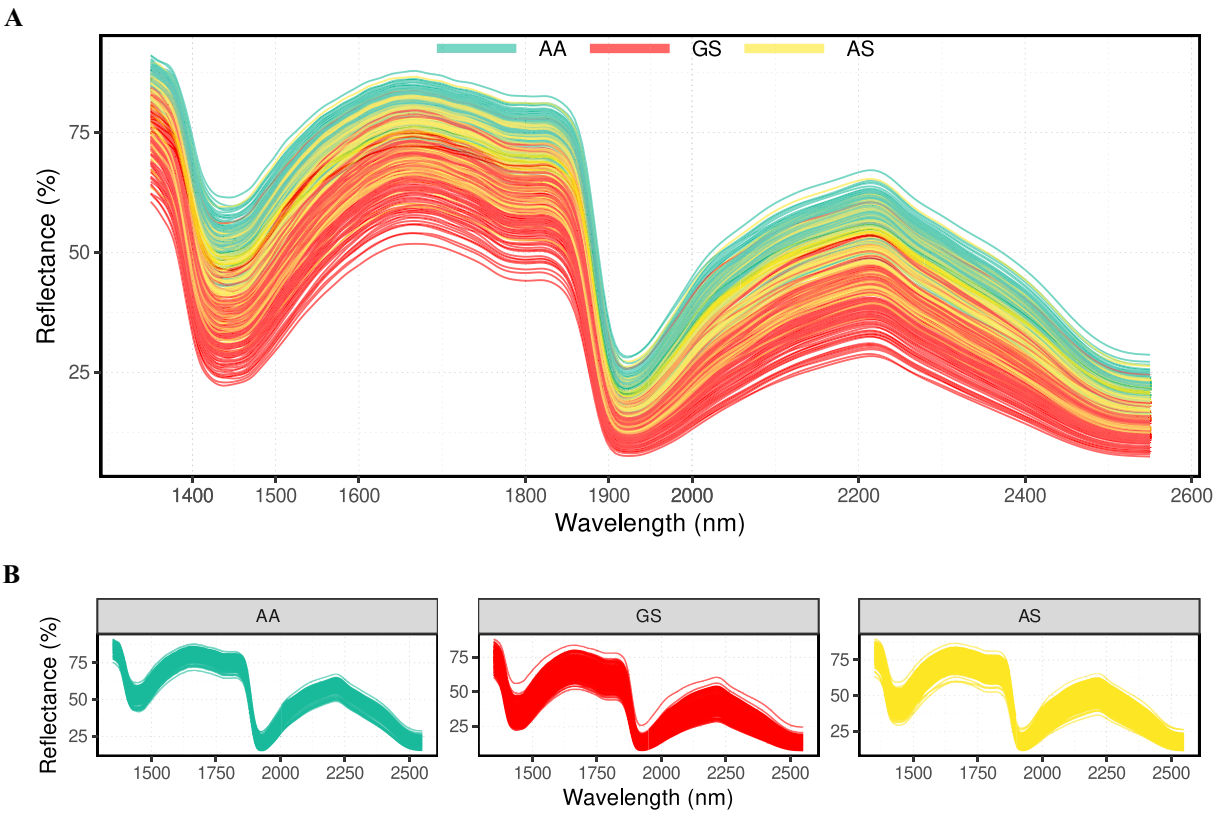

FIGURE 3
Near-infrared reflectance spectra of beech leaf tissues. **(A)** Overlay of all 449 raw NIR spectra (1348–2,551 nm) collected across symptom types, illustrating the overall spectral profiles. **(B)** Spectra separated by asymptomatic leaves (AA), asymptomatic tissue from symptomatic leaves (AS), and galled tissue from symptomatic leaves (GS).

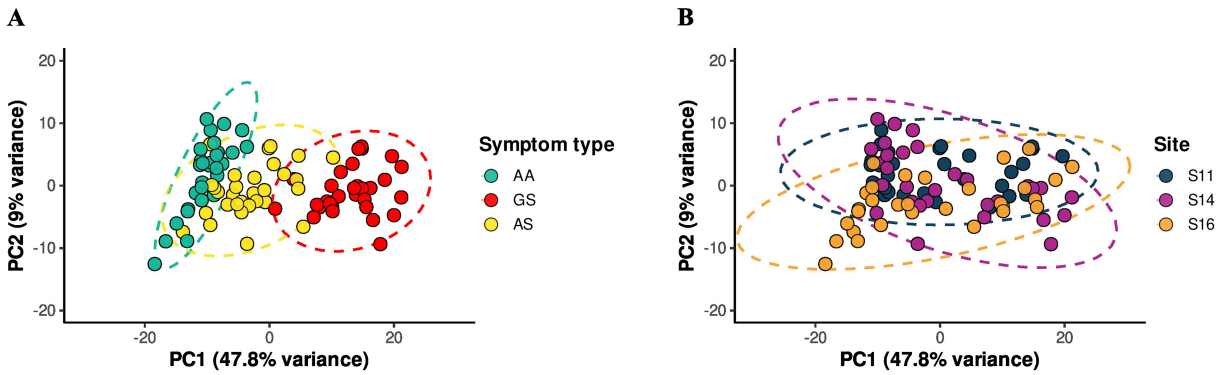

FIGURE 4
Principal component analysis (PCA) of NIR spectra. **(A)** PCA scores colored by symptom type: asymptomatic areas of asymptomatic leaves (AA), asymptomatic areas of symptomatic leaves (AS), and galled areas of symptomatic leaves (GS). **(B)** PCA scores colored by sampling site (S11, S14, S16). Ellipses indicate 95% confidence intervals for each group. The percentage of variance explained by each component is shown in parentheses on the axes.

global sPLS-DA selected three components (keepX = 50, 20, and 20 variables) and showed strong discrimination between AA and GS tissues, with consistently high AUC values (0.96–1.00). In contrast, AS tissues were less distinctly classified (AUC = 0.52–0.85), supporting the interpretation that AS represents an intermediate spectral and biochemical state between AA and GS tissues.

LC abundance across symptom types

Our probe-based qPCR assay showed high analytical performance across plates, with calibration curves exhibiting strong linearity between $\log_{10}$ nematode number and quantification cycle [$R^2 = 0.9927 \pm 0.0012$ (SD)]. The limit of detection was one nematode, and no amplification occurred from naïve tissue or no-template controls. Amplification efficiency was 82.83 ± 1.62 (SD), corresponding to regression slopes of −3.87 ± 0.06 (SD), all within acceptable bounds for quantitative detection. Although efficiency was slightly below the ideal range for qPCR (90–110%), this likely reflects inhibitory effects associated with complex plant matrices rather than limitations of the assay deployed. These metrics indicate excellent assay reproducibility and minimal inter-plate variability. All 90 tissue samples yielded quantifiable results after ensuring reproducible amplification across triplicates. LC abundance for each sample was calculated using the corresponding standard curve equation (Supplementary Table S3). Because LC counts were strongly right-skewed, values were transformed as $\log_{10}$ (abundance+1) prior to statistical analysis.

LC abundance differed significantly among symptom types ($F_{2,87} = 87.48$, $p < 0.001$), accounting for 66.8% of the total variance (Figure 5). Post-hoc Tukey's HSD tests revealed that GS tissues contained substantially higher numbers of LC than either AA (mean difference on the $\log_{10}$ scale = 2.28, 95% CI: 1.85–2.71, $p < 0.001$) or AS tissues (mean difference on the $\log_{10}$ scale = 1.74, 95% CI: 1.31–2.17, $p < 0.001$). AS tissues also contained significantly higher LC loads than AA (mean difference on the $\log_{10}$ scale = 0.54, 95%, CI: 0.11–0.97 $p = 0.001$).

LC abundance varied strongly with tissue symptom type and, to a lesser extent, across sites. A two-way ANOVA showed significant effects of both site ($F_{2,81} = 9.00$, $p < 0.001$) and site x symptom type interaction ($F_{4,81} = 5.83$, $p < 0.001$). Linear regression analyses further revealed symptom type-specific relationships between LC abundance and site-level disease severity. No significant association was observed for AA tissues ($R^2 = 0.017$, $p = 0.497$), AS tissues had a weak positive relationship ($R^2 = 0.147$, $p = 0.036$), and GS tissues exhibited a weak negative relationship ($R^2 = 0.150$, $p = 0.035$), suggesting site-level disease severity is a poor predictor of local LC abundance within individual tissue symptom classes.

Sequencing output and data quality

A total of 179 samples were successfully sequenced, comprising 90 ITS libraries (30 samples per symptom type) and 89 16S libraries (30 AA, 30 AS, and 29 GS). All retained samples yielded over 66,000 high-quality reads each and were included in downstream analyses. To ensure unbiased comparisons of α-diversity, ITS samples were rarefied to an even depth of 160,755 sequences per sample. Rarefaction curves for ITS data indicated that sequencing depth was sufficient to capture fungal diversity and accurately represent community composition (Supplementary Figure S4). Similarly, 16S rarefaction curves showed consistent sequencing depth across samples, with most curves reaching a clear plateau (Supplementary Figure S5). However, one AS 16S sample from site S14 displayed an atypical rarefaction profile, characterized by a disproportionate increase in observed bacterial richness relative to sequencing depth. This sample was identified as an outlier and excluded from downstream 16S analyses. These samples were rarefied to an even depth of 4,668 sequences per sample for α-diversity calculations. This final depth reflects the minimum remaining microbial library size after the removal of non-target sequences (e.g., mitochondria and chloroplasts) and unassigned taxa (Archaea and Eukaryota), which are common in plant-tissue-derived libraries.

α-diversity of bacterial and fungal communities

α -diversity was evaluated across symptom types and sampling sites (Table 1). Within the bacterial community, AS tissues exhibited the highest richness, exceeding that of both AA and GS tissues ($p = 0.000$, Supplementary Table S4). GS tissues showed the lowest diversity and evenness, whereas AA and AS tissues displayed broadly similar values ($p < 0.01$, Supplementary Table S4). No significant differences were found at site level ($p > 0.05$, Table 1; Supplementary Table S4). Overall, these patterns indicate that bacterial α-diversity is structured primarily by symptom type rather than sampling site.

10.3389/ffgc.2026.1782331

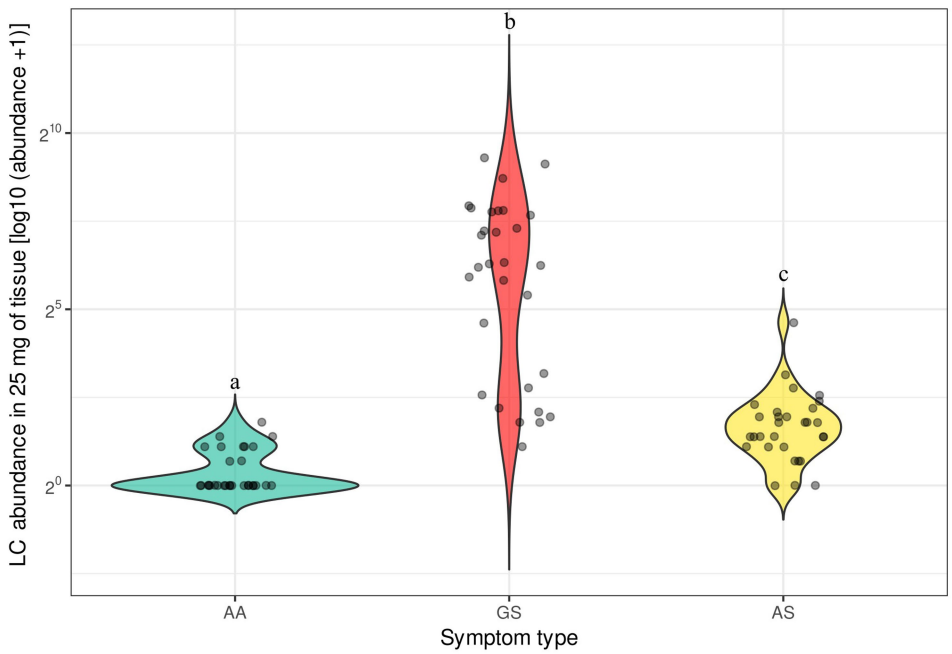

FIGURE 5
LC abundance in beech leaf tissue across symptom types. Violin plots show the distribution of LC $\log_{10}$ (abundance + 1) in asymptomatic areas of asymptomatic leaves (AA), asymptomatic areas of symptomatic leaves (AS), and galled areas of symptomatic leaves (GS). Letters indicate significant differences among groups (pairwise Tukey HSD $p < 0.05$) and points represent individual biological samples.

TABLE 1 α-diversity of beech bacterial and fungal leaf phytobiomes across symptom types and sampling sites.

16S rRNA bacterial community						
Symptom type	Observed ASVs	Chao	ACE	Shannon	Simpson	Pielou
AA	232.767 ± 83.738 a	247.091 ± 98.607 a	247.001 ± 98.722 a	4.376 ± 0.456 a	0.964 ± 0.023 a	0.811 ± 0.050 a
GS	195.517 ± 46.871 a	217.352 ± 55.781 a	218.044 ± 58.191 a	3.671 ± 0.553 b	0.926 ± 0.044 b	0.699 ± 0.090 b
AS	324.414 ± 82.250 b	361.798 ± 101.379 b	363.475 ± 105.088 b	4.611 ± 0.325 a	0.969 ± 0.012 c	0.801 ± 0.034 c
Site						
S11	263.759 ± 82.466 a	290.064 ± 104.082 a	291.843 ± 109.186 a	4.398 ± 0.441 a	0.967 ± 0.020 a	0.795 ± 0.055 a
S14	251.069 ± 103.930 a	275.059 ± 122.805 a	276.151 ± 123.800 a	4.253 ± 0.608 a	0.956 ± 0.038 a	0.779 ± 0.079 a
S16	237.700 ± 84.475 a	260.651 ± 94.092 a	260.075 ± 93.472 a	4.020 ± 0.682 a	0.938 ± 0.038 a	0.739 ± 0.093 a
ITS fungal community						
Symptom type	Observed ASVs	Chao	ACE	Shannon	Simpson	Pielou
AA	583.200 ± 130.052 a	584.181 ± 130.501 a	584.042 ± 130.379 a	3.014 ± 0.692 a	0.814 ± 0.121 a	0.472 ± 0.097 a
GS	484.267 ± 83.901 b	489.976 ± 84.832 b	489.403 ± 84.632 b	2.340 ± 0.423 b	0.761 ± 0.080 b	0.378 ± 0.061 b
AS	623.167 ± 175.599 a	627.752 ± 178.971 a	627.295 ± 177.729 a	2.804 ± 0.777 a	0.786 ± 0.126 ab	0.435 ± 0.108 a
Site						
S11	584.167 ± 127.654 a	587.008 ± 126.712 a	586.761 ± 126.758 a	2.951 ± 0.767 a	0.817 ± 0.102 a	0.462 ± 0.108 a
S14	569.000 ± 165.793 a	572.206 ± 168.588 a	571.571 ± 167.434 a	2.684 ± 0.646 a	0.781 ± 0.100 a	0.423 ± 0.088 a
S16	537.467 ± 142.863 a	542.695 ± 144.869 a	542.408 ± 144.521 a	2.523 ± 0.637 a	0.762 ± 0.128 a	0.401 ± 0.090 a

Mean ± SD values are shown for richness (Observed ASVs, Chao, ACE), diversity (Shannon, Simpson), and evenness (Pielou). Results are presented separately for bacterial (16S) and fungal (ITS) datasets and summarized by symptom type [asymptomatic areas of asymptomatic leaves (AA), asymptomatic areas of symptomatic leaves (AS), and galled areas of symptomatic leaves (GS)] and site (S11, S14, S16). Letters following statistics indicate significant differences among groups based on pairwise Wilcoxon rank-sum tests with Benjamini-Hochberg correction ($p < 0.05$). Both raw and adjusted p-values for each comparison are provided in Supplementary Table S4.

For the fungal community, richness estimates were highest in AS and AA tissues and significantly lower in GS tissues. Fungal diversity and evenness followed similar trends, with GS tissues exhibiting reduced Shannon diversity, Simpson diversity, and Pielou's evenness

compared with AA and AS tissues. No significant differences were found at site level ($p > 0.05$ Table 1; Supplementary Table S4). As with the bacterial community, fungal α-diversity was more strongly associated with symptom type than with sampling site, with GS tissues supporting the least diverse and least even fungal assemblages.

β-diversity of bacterial and fungal communities

PERMANOVA analyses showed that both symptom type and sampling site significantly structured foliar bacterial and fungal community composition ($p = 0.001$ for all terms). However, the magnitude of these effects differed between microbial groups. For the bacterial phytobiome, symptom type was the dominant explanatory factor (Supplementary Table S5; Figure 5), accounting for 26.6% of the variance in community composition, whereas sampling site explained 9.7% and the interaction between the two factors contributed an additional 13.0% (Supplementary Table S5). Pairwise comparisons among symptom types confirmed strong compositional divergence (Supplementary Table S6): AA and GS communities showed the greatest separation ($R^2 = 0.250$), followed by AA vs. $R^2 = 0.222$, with GS vs. AS also differing significantly but to a lesser extent ($R^2 = 0.150$).

In contrast, the fungal phytobiome was shaped predominantly by sampling site (Supplementary Table S5; Figure 6), which explained 21.8% of the observed variation in ITS community composition. Symptom type exerted a comparatively weak effect (4.8%, $p = 0.001$), and the interaction between symptom type and site explained an additional 4.2% of the variance ($p = 0.002$). Pairwise comparisons among sites (Supplementary Table S6) showed consistent effect sizes ($R^2 \approx 0.164$–0.178), indicating a clear site-level structuring of the foliar mycobiome.

Taxonomic composition of microbial communities

Across symptom types, both bacterial and fungal communities were dominated by a relatively small subset of genera, but their relative abundances differed markedly among AA, GS, and AS tissues (Figure 7). In the bacterial dataset, several of the most abundant genera varied significantly across symptom types ($p < 0.001$, Supplementary Tables S7, S8). 1,174–901-12 (Beijerinckiaceae) was most enriched in AA tissues (Lyu et al., 2024), whereas *Terriglobus* and P3OB-42 (Myxococcales) were most abundant in AS tissues (Šigutová et al., 2023; Zou et al., 2024). GS tissues showed high relative abundances of multiple genera, including *Wolbachia, Luteibacter, Pseudomonas, Pedobacter*, and the *Allorhizobium–Neorhizobium–Pararhizobium–Rhizobium* complex. These pronounced shifts demonstrate that bacterial community composition is strongly structured by symptom type.

Fungal communities displayed more homogeneous taxonomic profiles across symptom types, with *Ramularia* and *Erysiphe* together accounting for more than 60% of total relative abundance. Among the dominant genera, only *Trichomerum* and unidentified Helotiales were significantly enriched in AA tissues ($p < 0.001$, Supplementary Tables S7, S8). No other major fungal genera differed among symptom types, consistent with the stronger site-level rather than symptom type-level structuring observed in the fungal α-diversity analyses.

Indicator species analysis further supported these taxonomic patterns. Across the bacterial dataset, 93 genera were significantly associated with individual symptom types ($p < 0.05$), with an additional seven linked to combinations of types (Supplementary Table S9). Indicator values (IndVal) ranged from 0.179 to 0.576; strong associations were defined as $p < 0.05$ and IndVal > 0.450. AA tissues contained three indicator genera, with *Lysobacter* showing the strongest association (IndVal = 0.409, $p = 0.001$); GS tissues exhibited 24 indicator genera, among which *Wolbachia*, *Luteibacter*, and *Pseudomonas* had the highest indicator values (IndVal ≥ 0.515, $p = 0.001$); while AS tissues supported the largest indicator set, comprising 66 genera; those with the strongest associations (IndVal ≥ 0.45 and $p \leq 0.003$) included *Cutibacterium, Terriglobus, Acidiphilium, Quadrisphaera, Brevundimonas, Methylobacterium–Methylorubrum, Sphingomonas, Bdellovibrio, Amnibacterium, Frondihabitans, Limnobacter, P3OB-42, Phaselicystis, Bacillus, Hymenobacter,* 1,174–901-12, *Sphingoaurantiacus*, and *Acidipila–Silvibacterium*. Among

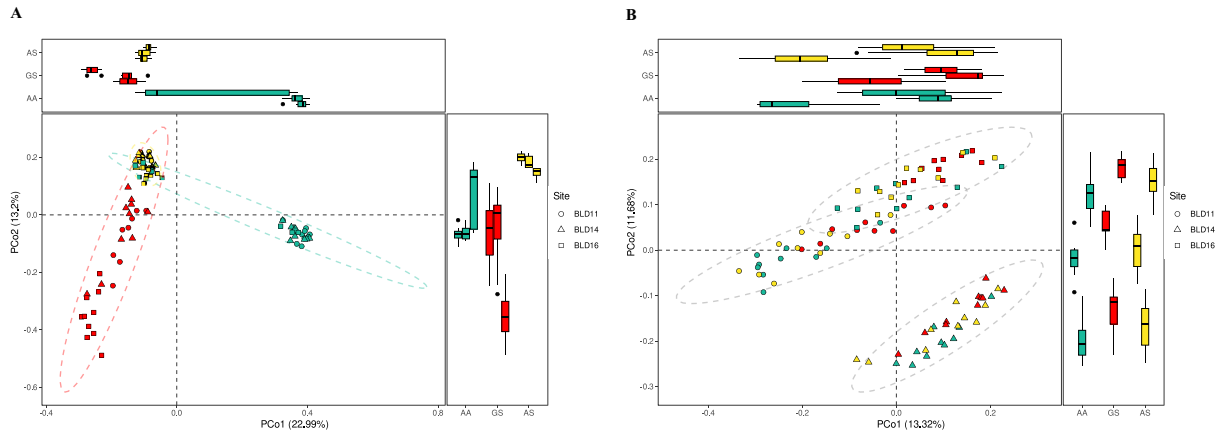

FIGURE 6
Principal coordinates analysis (PCoA) of bacterial **(A)** and fungal **(B)** communities based on Bray–Curtis dissimilarities. Samples are colored by symptom type [asymptomatic areas of asymptomatic leaves (AA), asymptomatic areas of symptomatic leaves (AS), and galled areas of symptomatic leaves (GS)] and shaped by sampling site (BLD11, BLD14, BLD16). Ellipses represent 95% confidence intervals. For bacterial communities **(A)**, ellipses are colored by symptom type whereas for fungal communities **(B)**, ellipses are shown by sampling site to reflect the primary structuring factor in each dataset. The percentage of variance explained by each axis is indicated in parentheses.

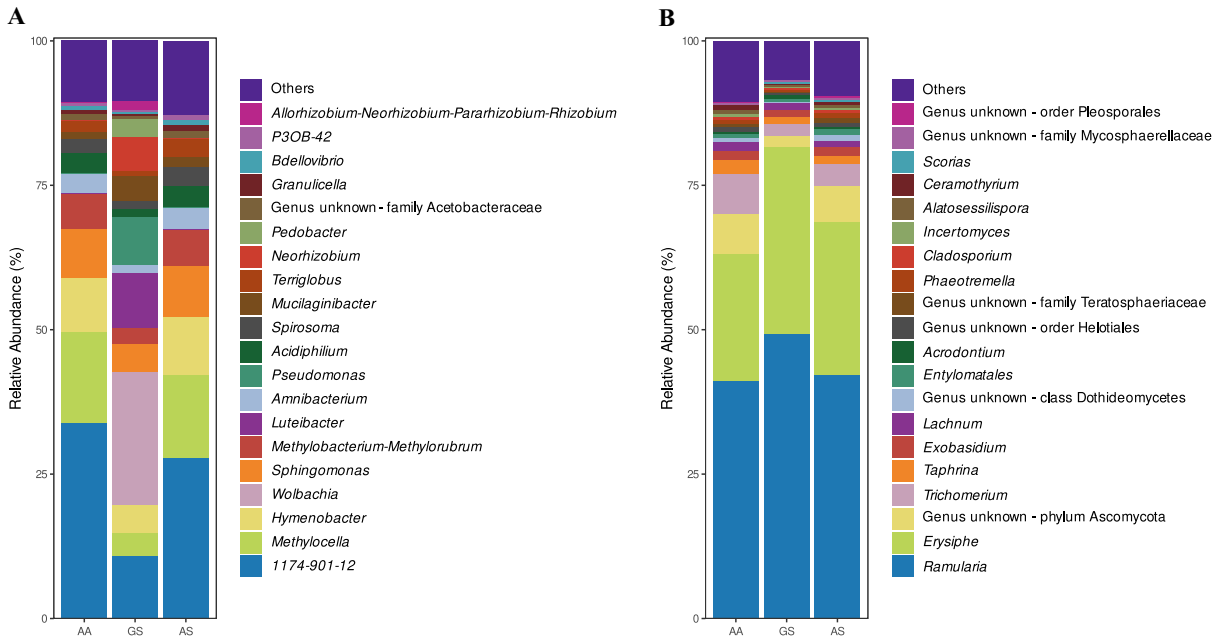

FIGURE 7
Composition of bacterial **(A)** and fungal **(B)** foliar communities by symptom type. Stacked barplots show the relative abundance of the 20 most dominant genera detected in asymptomatic areas of asymptomatic leaves (AA), asymptomatic areas of symptomatic leaves (AS), and galled areas of symptomatic leaves (GS). A similar analysis grouped by sampling site (S11, S14, S16) is presented in Supplementary Figure S5.

combinations of symptom types, the strongest association was observed for the AA + AS group, with *Methylocella* (IndVal = 0.546, $p = 0.001$).

In the fungal dataset, 59 genera were significantly associated with individual symptom types and 25 with combinations (Supplementary Table S10). IndVal values ranged from 0.184 to 0.429, with all indicator values below 0.45, indicating the absence of strong associations. AA tissues had 30 indicator genera, with *Mycena* showing the strongest association (IndVal = 0.429, $p = 0.002$); GS tissues had 14 indicator genera, with *Sclerostagonospora* exhibiting the highest indicator value (IndVal = 0.354, $p = 0.005$); while AS tissues contained 15 indicator genera, led by *Phaeotremella* (IndVal = 0.333, $p = 0.003$). For combinations, the AA + AS group showed the strongest association, driven by *Hyphodontia* (IndVal = 0.375, $p = 0.004$).

Phytobiome–LC abundance association

db-RDA revealed a significant association between LC abundance and bacterial community composition. In the unconstrained model, LC abundance explained 11.0% of the variation in bacterial β-diversity ($F_{1,86} = 10.60$, $p = 0.001$), indicating a clear compositional shift along the LC abundance gradient (Figure 8). Because LC abundance covaries with symptom type, partial db-RDA was used to quantify its independent effect. After conditioning on symptom type, LC abundance remained a statistically significant but very modest predictor ($F_{1.84} = 3.82$, $p = 0.001$), explaining only 3.2% of the residual variation (Figure 8). These results demonstrate that LC contributes both shared and symptom-independent components of bacterial community variation in beech leaves.

In the fungal dataset, LC abundance was also associated with community turnover, but with smaller effect sizes. The full model explained 4.0% of Bray–Curtis dissimilarity ($F_{1,88} = 3.69$, $p = 0.001$), and this association persisted after conditioning on symptom type (2.5% explained variance; $F_{1,86} = 2.28$, $p = 0.003$). Together, these patterns confirm that bacterial communities are more tightly linked to symptom type and LC abundance, whereas fungal assemblages exhibit comparatively weaker and more spatially structured responses.

Taxon-level associations between the bacterial community and LC abundance revealed several ASVs with significant positive and negative relationships (Figure 9). Several bacterial ASVs were positively associated with increasing LC abundance, including *Pseudomonas* (ASV68; $\beta = 0.89$, $q = 0.0031$), *Wolbachia* (ASV80; $\beta = 0.76$, $q = 0.0105$), *Pedobacter* (ASV62; $\beta = 0.54$, $q = 0.0478$), a second *Wolbachia* ASV (ASV112; $\beta = 0.38$, $q = 0.0015$), and a second *Pseudomonas* ASV (ASV394; $\beta = 0.46$, $q = 0.0157$). In contrast, negative associations with LC abundance were only significant for one ASV, *Wolbachia* (ASV27; $\beta = -1.04$, $q = 0.0021$).

NIR–LC association analysis

Rather than classifying symptom type, NIR spectral reflectance was evaluated for its ability to predict continuous variation in LC abundance. PCA of NIR spectra reproduced the symptom-based separation described above, and overlaying LC abundance revealed a clear intensity gradient along PC1, corresponding to increasing log-transformed nematode abundance (Supplementary Figure S7). Spearman's rank correlations indicated widespread and strong associations between spectral reflectance and LC abundance, with 200 of 246 wavelengths (81.3%) showing significant correlations ($q < 0.05$). The strongest association occurred at 1406.9 nm ($|\rho| = 0.794$, $q = 2.3 \times 10^{-18}$), with adjacent peaks spanning 1,401–1,404 nm. Additional regions of high correlation were observed at 1,370–1,373 nm, 1,958–1,964 nm, and 2,481–2,490 nm.

The PLS-R model revealed strong predictive relationships between NIR spectra and log-transformed LC abundance, with the optimal model explaining 89% of the variance ($R^2 = 0.89$) and a RMSEP of

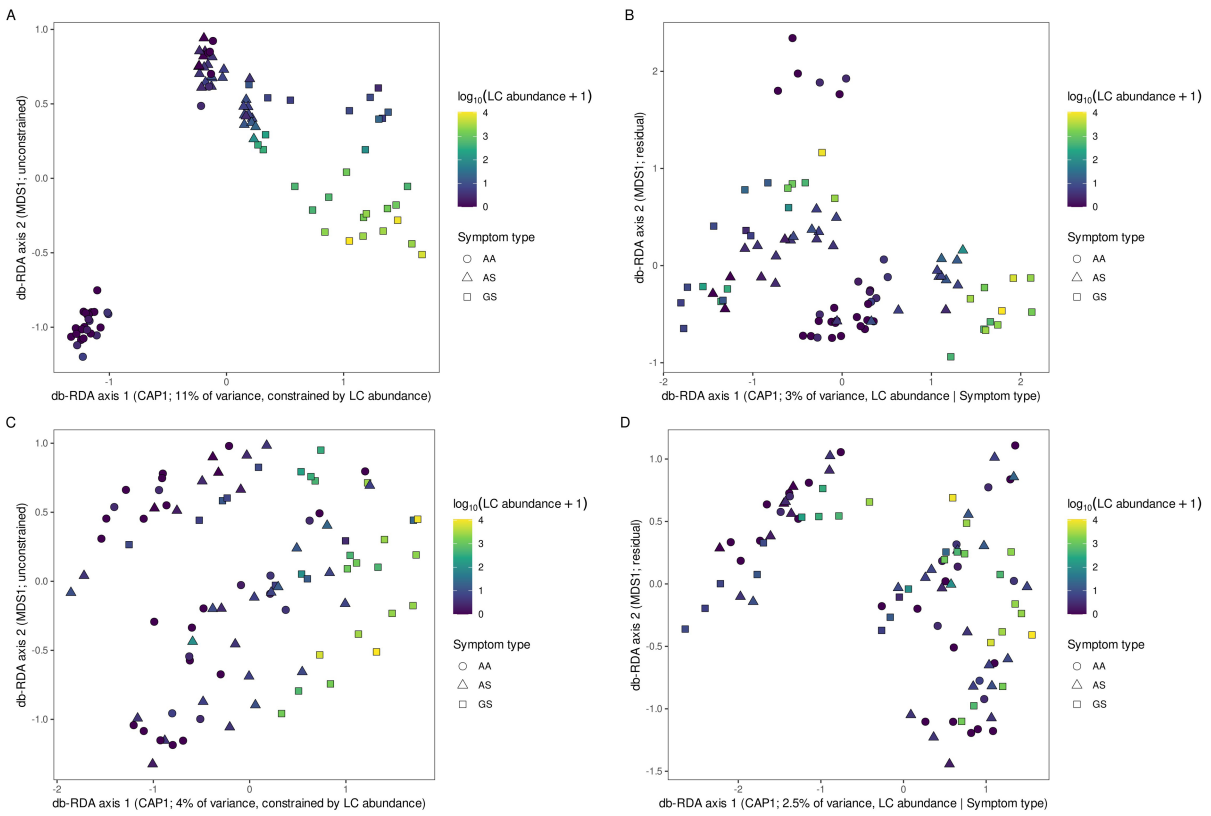

FIGURE 8
Constrained and partial distance-based redundancy analyses (db-RDA) illustrating the effects of LC abundance on foliar bacterial and fungal community composition. **(A,B)** Full and partial db-RDA models for the 16S dataset, and **(C,D)** corresponding models for the ITS dataset. In all panels, points represent individual samples, are shaped by Symptom type, and are colored according to $\log_{10}$ (LC abundance + 1).

1.462 with seven components. Models based on untransformed counts performed substantially worse (R^2 = 0.204; RMSEP = 1561.175), indicating that LC abundance was more accurately captured on a logarithmic scale. Cross-validated models retained high predictive accuracy (R^2 = 0.67 ± 0.16; RMSEP 1.65), with the optimal model containing two components. Permutation testing confirmed that the model performance exceeded that expected by chance (observed R^2 = 0.896; $p < 0.01$). Additional performance metrics (MAE = 0.74; Bias ≈ 0; RPD = 3.12) further supported the robustness of NIR-based prediction of nematode abundance.

NIR and 16S prediction

Phytobiome prediction models based on NIR were developed independently of symptom severity and the LC abundance models described above. While host spectral traits may reflect both pathogen load and BLD microbial shifts, the phytobiome prediction analysis was conducted using only the bacterial community ordination axes derived from the 16S community composition. Thus, spectral models were trained to reconstruct multivariate bacterial community composition rather than disease-related measurements. NIR spectral data also showed strong capacity to predict bacterial community composition derived from 16S rRNA gene sequencing (Table 2). Across all models evaluated, NIR-based predictions achieved high accuracy, indicating that variation in microbial community structure is reflected in host spectral signatures. Nonlinear models consistently outperformed linear approaches, with SVM and random forest models exhibiting the highest predictive accuracy and lowest calibration error, whereas elastic net regression showed comparatively weaker performance.

Joint prediction of the first two PCoA axes enabled direct reconstruction of bacterial community configuration in ordination space (Figure 10). Both random forest and SVM models accurately reproduced primary and secondary gradients of community differentiation, with moderately high joint R^2 values (0.791 and 0.801, respectively, Table 2), and slightly stronger performance for SVM. These results demonstrate that NIR spectra encode multivariate information sufficient to reconstruct bacterial β-diversity patterns beyond a single dominant axis.

Discussion

In this study, we provide estimates of LC load in tissues of variable symptom type (based on an in-house qPCR assay -Torres-Bedoya et al., in preparation), and gain insights, through NIR spectroscopy, into shifts in host biochemistry and phytobiomes associated with BLD. Specifically, our spectral and multivariate analyses show that variation in leaf chemical traits during BLD is structured primarily by symptom type rather than geographic location. This pattern was expected, given the conspicuous visual interveinal banding and gall formation characteristic of BLD symptomatic leaves. AA and GS tissues (Figure 1) exhibited the greatest biochemical divergence;

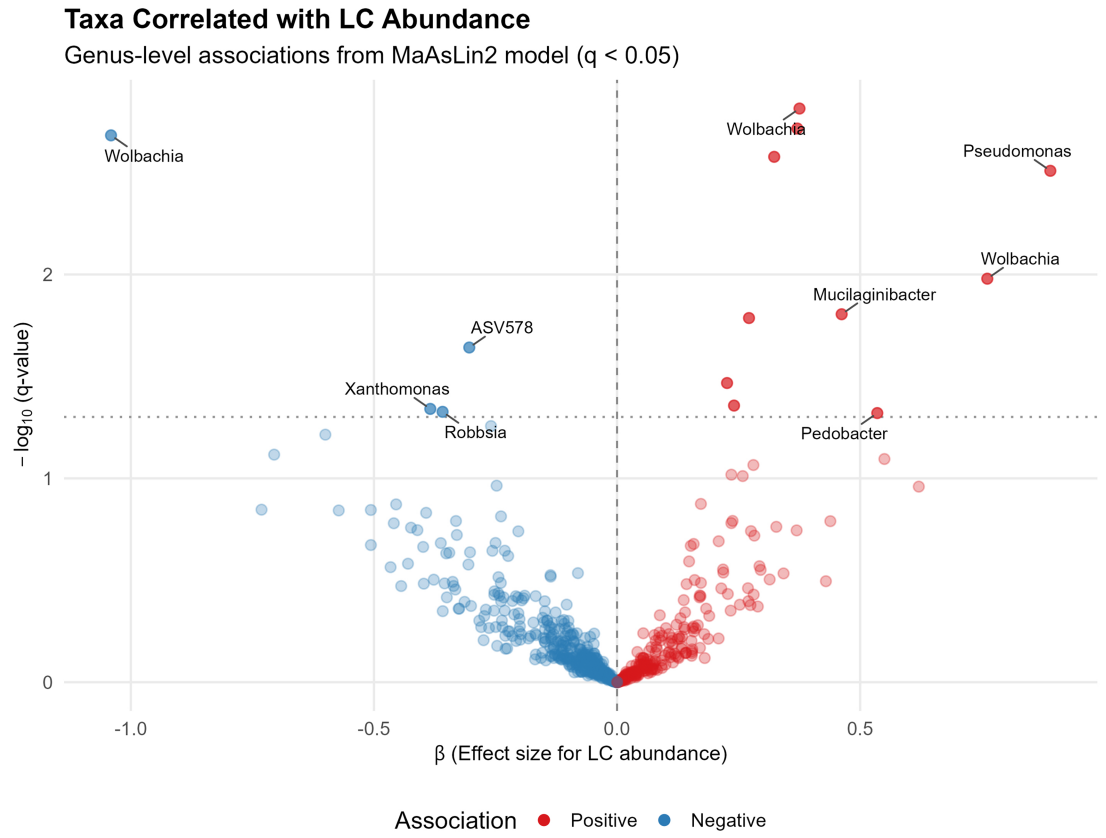

FIGURE 9
Volcano plot showing ASV-level associations between bacterial taxa and LC abundance. Each point represents an ASV taxon tested in the MaASLin2 model, with the *x*-axis showing the estimated effect size (β) and the *y*-axis the corresponding significance (−log10 *q* value). Red indicates positive relationships, while blue indicates negative relationships with LC abundance. Labeled points represent the top five taxa showing the strongest positive and negative associations.

TABLE 2 Predictive performance of machine learning models using NIR spectral profiles to estimate 16S phytobiome community composition (PCoA1 axis or PCoA1 + PCoA2).

Parameters for machine learning models				
Model	R^2	RMSE	ACE	Parameters
PCoA 1				
Random forest	0.879	0.097	73.3	Ntree = 500
Support vector machine	0.901	0.12	102.81	Kernel = radial, cost = 10, gamma = 0.01, CV fold = 5
Elastic net	0.746	0.127	97.925	Alpha = 0.5, ʎmin = 0.0247
PLSR	0.862	0.095	77.641	Ncomp = 10 (validated, CV RMSEP = 0.0999)
PCoA 1+PCoA 2				
Random forest	0.791	0.154	137.52	Ntree = 500
Support vector machine	0.801	0.179	161.42	Kernel = radial, cost = 10, gamma = 0.01, CV fold = 5

Models were trained to predict the PCoA axes of the Bray–Curtis dissimilarity matrix. Across models, NIR-based prediction reconstructed major gradients of bacterial community differentiation, with the highest performance observed for support vector machine and random forest approaches. Performance metrics are reported: the coefficient of determination (R^2), Root Mean Square Error (RMSE), scaled Average Calibration Error (ACE), and model parameters.

in contrast, AS tissues consistently occupied an intermediate position, indicating a gradient of physiological alteration (Figure 4). GS regions correspond to zones of high nematode aggregation and are characterized by elevated chlorophyll content (Miles, 2023; Vieira et al., 2023).

The wavelengths contributing most to spectral differentiation between AA and GS tissues (Figures 3, 4) were concentrated in regions of the NIR spectrum associated with leaf water status, structural carbohydrates, and mixed chemical bonds (i.e., C-H and C-N) (Türker-Kaya and Huck, 2017). The enrichment of these spectral features in

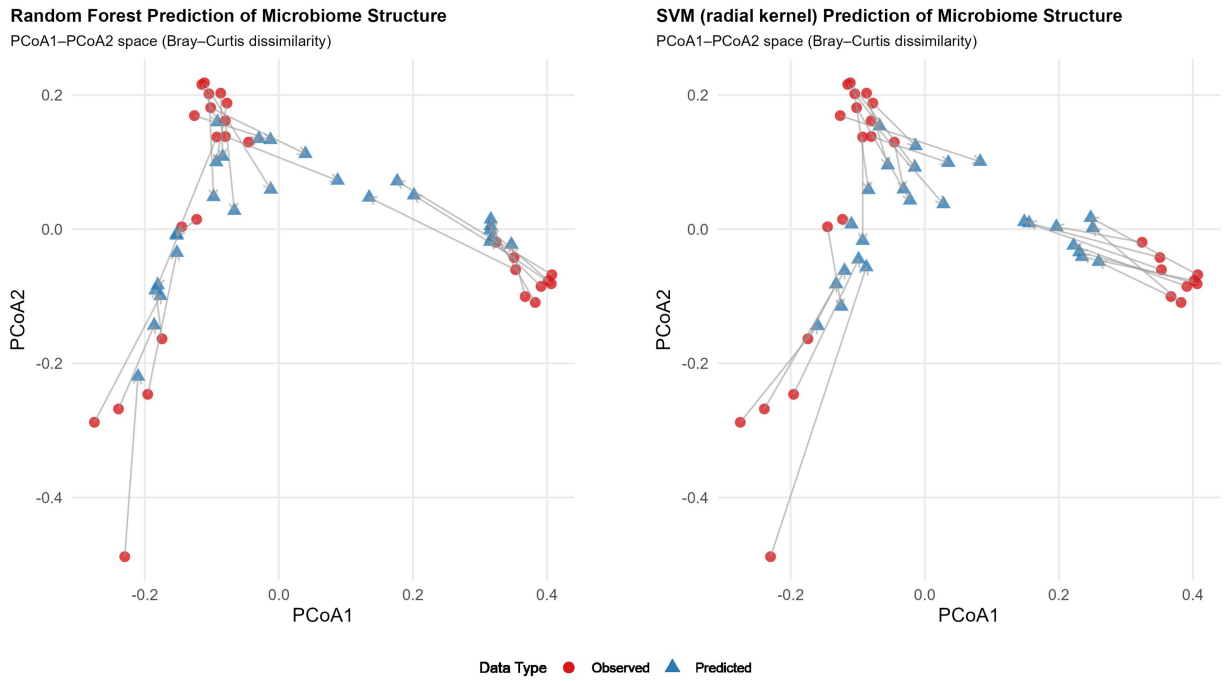

FIGURE 10
Prediction of bacterial microbiome community composition from NIR spectra. Left panel: Random Forest predictions of bacterial community composition in PCoA-PCoA2 space derived from Bray-Curtis dissimilarities. Right panel: Support Vector Machine (radial kernel) forecasts of the same ordination space. In both panels, red circles represent observed microbiome ordination coordinates, and blue triangles represent NIR-predicted coordinates. The gray line segments connect observed and predicted positions for each sample.

GS tissues is consistent with our hypothesis of localized increases in water content and nitrogen accumulation in regions expressing disease symptoms. Although nematode infections are often associated with reduced whole-plant water status, nematode-induced feeding sites and gall tissues function as metabolically active sink structures exhibiting enhanced capacity for solute and water import from host vascular tissues (Baldacci-Cresp et al., 2015; Grundler and Hofmann, 2011; Petrikovszki et al., 2023; Zhao et al., 2018), a phenomenon similar to what is known in plant bacteriology as "pathogen niche establishment" (Roussin-Léveillée et al., 2024).

LC abundance varied strongly among symptom types, with GS tissues consistently supporting higher nematode abundance than AA and AS tissues (Figure 5). Notably, GS tissues exhibited substantial variability in LC abundance. We posit that, because initial damage occurs in buds, wide variation in nematode abundance within buds can produce comparable banding symptoms. Additionally, leaves varied in the number of banded regions, which may reflect differences in the initial nematode populations present within the buds.

LC abundance increased significantly across symptom types (AA < AS < GS), with both AA and AS tissues harboring substantially lower LC densities than GS tissues. The low LC loads observed in AA tissues may indicate detections that do not necessarily reflect active infection from the bud, but rather superficial contamination due to horizontal transfer in wind, rain splash, or other passive transport (Goraya et al., 2024), which may account for the low frequencies. Interestingly, for AS tissues, our results suggest that once nematodes colonize interveinal sectors of developing leaf primordia within the bud, they remain confined to those sectors until they emerge, and then migrate to new buds in the fall, i.e., after our late-summer sampling period. However, the timing and mechanisms governing LC migration from leaf tissues to overwintering buds remain unknown, representing a gap in understanding LC life history and disease cycle.

The phytobiome predictive models developed in this work demonstrate that NIR spectral profiles capture meaningful variation in bacterial community composition within beech leaves, supporting our hypothesis that underlying host phytobiome states are reflected in chemical and, therefore, spectral traits. This suggests that phytobiome assembly and transitions from eubiosis to dysbiosis can be inferred non-destructively from spectral data, enabling rapid phenotyping of pathogen-associated microbial shifts before visible symptoms appear (Figure 10; Table 2). Because spectral inference, unlike molecular assays, is not targeted to specific taxa, it might also detect emergent or uncharacterized disease states in a biosurveillance context, before PCR-based confirmation.

Model evaluation progressed from predicting the primary bacterial ordination axis (PCoA1) to jointly predicting PCoA1 and PCoA2 using RF and SVM approaches. As expected, predictive performance declined with increasing dimensionality, likely reflecting added complexity and within-host heterogeneity captured by pooled tissues (Kusonmano et al., 2012). Nevertheless, strong correspondence between observed and predicted ordination clustering patterns indicates that NIR spectra retained key features of bacterial community composition.

Extending this modeling to LC abundance, PLS-R showed substantially improved performance after log-transformation of LC counts, consistent with the right-skewed distribution of LC abundances, achieving a high predictive capacity ($R^2 = 0.67$). While further refinement will require larger and more diverse datasets, these results support the deployment of NIR spectroscopy as a scalable tool for identifying elevated infection pressure and monitoring disease responses non-destructively, positioning it as both an early-detection

and surveillance resource for beech management (Crandall et al., 2020; Fearer et al., 2022a; Galvan et al., 2023; Gold et al., 2020). As this study sampled asymptomatic leaves from trees already exhibiting BLD symptoms, our results provide initial validation that symptomatic and asymptomatic tissues exhibit distinct NIR spectral profiles. Future work should evaluate whether these spectral signatures can be used to detect infection in buds, and in trees or stands without visible symptoms, providing an opportunity for external validation. Incorporating naïve tissues into the model will also be important for establishing a true baseline and mitigating potential effects of systemic host responses associated with infection.

α-diversity patterns further support the interpretation that BLD is associated with localized disruption of foliar phytobiome structure. GS tissues exhibited the lowest diversity and evenness of bacterial communities, and in the fungal community also showed reduced richness, consistent with phytobiome simplification and a transition toward dysbiosis (Arnault et al., 2023; Torres-Bedoya et al., under review). In contrast, AA and AS tissues maintained comparable fungal community diversity, with finer tissue resolution revealing stratification by symptom within symptomatic leaves. These results indicate that LC feeding zones are consistently associated with reduced phytobiome diversity, whereas adjacent AS regions often maintain diversity similar to or higher than that of AA tissues, reflecting localized microbial reorganization. Although the underlying drivers of these shifts cannot be resolved here, their structured nature highlights a close coupling between host physiological state and phytobiome assembly. These α-diversity patterns support the premise that host spectral phenotypes reflect physiological states that covary with phytobiome structure, indicating that dysbiosis-associated diversity loss is biologically meaningful and detectable via NIR.

On the other hand, β-diversity patterns revealed contrasting drivers of community turnover between bacterial and fungal assemblages (Supplementary Tables S5, S6). Bacterial communities were structured primarily by symptom type, with clear separation among AA, AS, and GS tissues, but most pronounced between AA and GS, consistent with localized dysbiosis in LC-enriched regions. In contrast, fungal communities were driven mainly by geographic location rather than symptom type, indicating that foliar mycobiome assembly is more strongly shaped by broader environmental, climatic, or site-specific ecological filters than by within-leaf disease state. Similarly, while LC abundance was a strong driver of foliar bacterial community composition, it exerted a comparatively weaker influence on fungal assemblages. Bacterial communities shifted measurably along the LC gradient, and this effect remained significant after accounting for symptom type, indicating that LC contributes both shared and symptom-independent components of bacterial community variation. In contrast, fungal communities showed only modest associations with LC abundance, which were further reduced after controlling for symptom type, reinforcing the role of broader site-level filters in structuring the mycobiome.

The genus *Wolbachia* was a significant indicator of GS tissues and was positively associated with nematode abundance in our linear modeling (Figure 9; Supplementary Table S9), confirming previous reports (Burke et al., 2020, 2024; Ewing et al., 2021; Miles, 2023). Although *Wolbachia* is broadly associated with insects and several plant-parasitic nematodes (Fenn et al., 2006), the recurrent detection of this genus in BLD-affected foliage supports the hypothesis that it is most likely nematode-associated in this system, and at least part of the apparent foliar "dysbiosis" may derive from the nematode's own microbiome rather than solely from restructuring of the native beech phytobiome. Targeted characterization of the LC microbiome is therefore critical to determine whether *Wolbachia* represents a stable nematode-associated symbiont and to clarify its functional and phylogenetic placement within the BLD system. Not all *Wolbachia* exhibited uniform responses; one ASV was negatively associated with LC abundance, suggesting heterogeneous responses within the genus, alternative hosts (e.g., *Aphelenchoides* spp., which has also been associated with forest environments and beech tissues), or differential compatibility among strains in LC (Brown et al., 2018; Goraya et al., 2024; Wasala et al., 2019, 2023; Christopher Taylor, personal communications). Although biocontrol strategies targeting *Wolbachia* have not yet been investigated in plant-pathogenic nematode systems, the concept has been explored theoretically for filarial nematodes, and *Wolbachia*-based management approaches have been successfully implemented in insect systems, such as mosquitoes (Gong et al., 2023; Johnston and Taylor, 2007; Slatko et al., 2010).

Of additional interest was the bacterial genus *Pseudomonas*, which was also detected in prior work on BLD (e.g., Ewing et al., 2021); however, its positive association with LC must be interpreted cautiously, given the genus diversity (Silby et al., 2011). While some *Pseudomonas* spp. respond to treatments targeted specifically at nematodes and are enriched in nematode-infected tissues (Kaplan et al., 2009; Kunda et al., 2024; Leontopoulos et al., 2017; Yergaliyev et al., 2020), our data cannot confirm whether this association reflects a functional role in disease progression, a host-mediated response to infection, or broader phytobiome restructuring driven indirectly by LC.

Beyond *Wolbachia* and *Pseudomonas*, several additional bacterial taxa showed significant associations with LC abundance, including ASVs assigned to *Pedobacter*, *Xanthomonas*, and *Robbsia*. These taxa are ecologically diverse and can function as opportunists, commensals, or beneficial associates depending on host condition and environmental context (Figueiredo et al., 2025; Mannaa et al., 2018; Thomas et al., 2024; Timilsina et al., 2020). Accordingly, their enrichment or depletion within symptomatic tissues cannot be unambiguously attributed to direct interactions with LC. These shifts might reflect broader restructuring of the foliar phytobiome under dysbiotic conditions, potentially driven by altered resource availability, host physiological responses, or indirect biotic interactions. Resolving whether such taxa contribute functionally to disease progression, represent opportunistic colonizers of LC-modified tissue, or participate in host compensatory responses, will require targeted studies. Thus, while these associations are biologically interesting and meaningful, their mechanistic interpretation remains open and needs further investigation.

Although this study focused on the contrasts among symptom types, both between and within leaves, extending this framework to include naïve tissues, buds and leaves spanning higher disease severity classes would further elucidate how metabolic shifts and LC abundance vary with disease. In particular, leaves exhibiting crinkling and necrosis from LC infection may display altered fungal colonization dynamics and potentially stronger pathogen-fungal associations than we have found in this study (Fearer, 2022; Vieira et al., 2023). Moreover, since foliar chemistry, spectral profiles, and LC loads may vary seasonally, incorporating additional

sampling time points across the growing season will be critical for further contextualizing these patterns. Importantly, future model development will require one-to-one pairing of an NIR spectrum with phytobiome data derived from the same spatially explicit tissue regions, rather than pooled or adjacent samples, to ensure that spectral signatures are directly linked to the microbiome composition they are intended to predict. Such spatially matched sampling is essential for improving predictive accuracy, reducing the effect of within-leaf heterogeneity, and strengthening the biological interpretability of the model. Nevertheless, the results herein support the validity of NIR spectroscopy as a rapid, non-destructive tool for detecting phytobiome dysbiosis and quantifying pathogen load in BLD, with broad applicability to other plant-pathogen systems.

Conclusion

Overall, this research provides new insight into American beech-LC interactions and establishes a foundation for future targeted studies of the LC-associated phytobiome, metabolite profiling, and in-depth molecular characterization of host leaf responses. Furthermore, it supports the development of NIR-phytobiome models both within and beyond the context of BLD, suggesting that integrative frameworks linking pathogen load, host-derived physiological traits, and phytobiome assembly are feasible in this and potentially other pathosystems.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.5061/dryad.m37pvmdgv.

Author contributions

AM: Visualization, Data curation, Investigation, Validation, Conceptualization, Formal analysis, Writing – review & editing, Writing – original draft, Methodology. ET-B: Visualization, Validation, Formal analysis, Writing – original draft, Investigation, Data curation, Writing – review & editing, Methodology, Conceptualization. AC: Methodology, Writing – review & editing, Conceptualization, Funding acquisition, Resources. PB: Methodology, Supervision, Writing – review & editing, Project administration, Conceptualization, Resources, Funding acquisition.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This project was funded, in part, by USDA Forest Service Joint Venture Agreement 24-JV-11242314-012 with Ohio State University. Additional funding included grants to AM from the Kurt Gottschalk Science Fund (Society of American Foresters) and the Infectious Disease Institute Host Defense and Microbial Biology New Techniques Grant Program (Ohio State).

Acknowledgments

We thank Justina Ker for her assistance with sample processing and nematode quantification. We also thank Soumya Gosh, Selu Adams, and Katie Decamp for their assistance with field data collection, and the staff of the Allegheny National Forest for granting us access to the field sites.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author PB declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. Microsoft Copilot was used to rewrite selected passages to improve clarity.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/ffgc.2026.1782331/full#supplementary-material

References

Albright, T. A., Butler, B. J., Caputo, J., et al. (2023). "Pennsylvania forests 2019: summary report," in *Resour. Bull. NRS-131*, (Madison: US Department of Agriculture, Forest Service, Northern Research Station), 1–36.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed November 2, 2025).

Arnault, G., Mony, C., and Vandenkoornhuyse, P. (2023). Plant microbiota dysbiosis and the Anna Karenina principle. *Trends Plant Sci.* 28, 18–30. doi: 10.1016/j.tplants.2022.08.012

Arriola, Í. A., and Isaias, R. M. d. S. (2021). Extending the knowledge on the histological patterns of leaf galls induced by *Ditylenchus gallaeformans* (Nematoda) on *Miconia* (Melastomataceae) hosts. *Flora* 274:151753. doi: 10.1016/j.flora.2020.151753

Baldacci-Cresp, F., Maucourt, M., Deborde, C., Pierre, O., Moing, A., Brouquisse, R., et al. (2015). Maturation of nematode-induced galls in *Medicago truncatula* is related to water status and primary metabolism modifications. *Plant Sci.* 232, 77–85. doi: 10.1016/j.plantsci.2014.12.019

Balla, A., Silini, A., Cherif-Silini, H., Chenari Bouket, A., Moser, W. K., Nowakowska, J. A., et al. (2021). The threat of pests and pathogens and the potential for biological control in forest ecosystems. *Forests* 12:1579. doi: 10.3390/f12111579

Boateng, E. Y., Otoo, J., and Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. *J. Data Anal. Inf. Process.* 8, 341–357. doi: 10.4236/jdaip.2020.84020

Bonello, P., Campbell, F. T., Cipollini, D., Conrad, A. O., Farinas, C., Gandhi, K. J. K., et al. (2020). Invasive tree pests devastate ecosystems—a proposed new response framework. *Front. Forests Glob. Change* 3:2. doi: 10.3389/ffgc.2020.00002

Bonello, P., Conrad, A. O., Sadiković, D., Liziniewicz, M., and Cleary, M. (2025). Point-of-care diagnostics and resistance phenotyping to combat ash dieback. *Front. Forests Glob. Change* 8:1588428. doi: 10.3389/ffgc.2025.1588428

Breiman, L. (2002). *Manual on Setting up, Using, and Understanding Random Forests v3.1*, vol. 1 Berkeley, CA: Statistics Department University of California, 3–42.

Brown, A. M. V., Wasala, S. K., Howe, D. K., Peetz, A. B., Zasada, I. A., and Denver, D. R. (2018). Comparative genomics of *Wolbachia–Cardinium* dual endosymbiosis in a plant-parasitic nematode. *Front. Microbiol.* 9:2482. doi: 10.3389/fmicb.2018.02482

Burke, D. J., Carrino-Kyker, S. R., Hoke, A. J., Galloway, E., Martin, D., and Chick, L. (2024). Effects of the nematode Litylenchus Crenatae Subsp. Mccannii and beech leaf disease on leaf fungal and bacterial communities on *Fagus Grandifolia* (American beech). *Appl. Environ. Microbiol.* 90, e0014224–e0014224. doi: 10.1128/aem.00142-24

Burke, D. J., Hoke, A. J., and Koch, J. (2020). The emergence of beech leaf disease in Ohio: probing the plant microbiome in search of the cause. *For. Pathol.* 50:e12579. doi: 10.1111/efp.12579

Cale, J. A., Garrison-Johnston, M. T., Teale, S. A., and Castello, J. D. (2017). Beech bark disease in North America: over a century of research revisited. *For. Ecol. Manag.* 394, 86–103. doi: 10.1016/j.foreco.2017.03.031

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. doi: 10.1038/nmeth.3869

Carta, L. K., Handoo, Z. A., Li, S., Kantor, M., Bauchan, G., McCann, D., et al. (2020). Beech leaf disease symptoms caused by newly recognized nematode subspecies *Litylenchus crenatae mccannii* (Anguinata) described from *Fagus grandifolia* in North America. *For. Pathol.* 50:e12580. doi: 10.1111/efp.12580

Colbert-Pitts, M., Kantor, M. R., Jansen, A., Burke, D. J., and Vieira, P. (2025). Cellular dynamics of beech leaf disease on *Fagus sylvatica*. *Plant Pathol.* 74, 1389–1406.

Collin, A., Messier, C., Kembel, S., and Bélanger, N. (2017). Low light availability associated with American beech is the main factor for reduced sugar maple seedling survival and growth rates in a hardwood forest of southern Quebec. *Forests* 8:413. doi: 10.3390/f8110413

Conrad, A. O., and Bonello, P. (2016). Application of infrared and Raman spectroscopy for the identification of disease resistant trees. *Front. Plant Sci.* 6:1152. doi: 10.3389/fpls.2015.01152

Conrad, A. O., Li, W., Lee, D.-Y., Wang, G.-L., Rodriguez-Saona, L., and Bonello, P. (2020). Machine learning-based presymptomatic detection of rice sheath blight using spectral profiles. *Plant Phenomics* 2020:8954085. doi: 10.34133/2020/8954085

Crandall, S. G., Gold, K. M., Jiménez-Gasco, M. D. M., Filgueiras, C. C., and Willett, D. S. (2020). A multi-omics approach to solving problems in plant disease ecology. *PLoS One* 15:e0237975. doi: 10.1371/journal.pone.0237975

De Cáceres, M., and Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology* 90, 3566–3574. doi: 10.1890/08-1823.1

De Freitas Silva, M., Degaspari, G. J., De Souza Fernandes, D. P., Azevedo, A. A., De Oliveira, C. M. G., and Buonicontro, D. S. (2025). Histopathological study of soybean (*Glycine max* (L.) Merr) infected by *Aphelenchoides pseudobesseyi* (Nematoda: Aphelenchoididae). *J. Plant Dis. Prot.* 132:177. doi: 10.1007/s41348-025-01176-0

Denk, T., and Grimm, G. W. (2009). The biogeographic history of beech trees. *Rev. Palaeobot. Palynol.* 158, 83–100. doi: 10.1016/j.revpalbo.2009.08.007

Du, K.-L., Jiang, B., Lu, J., Hua, J., and Swamy, M. N. S. (2024). Exploring kernel machines and support vector machines: principles, techniques, and future directions. *Mathematics* 12:3935. doi: 10.3390/math12243935

Esposito Vinzi, V., and Russolillo, G. (2013). Partial least squares algorithms and methods. *WIREs Comput. Stat.* 5, 1–19. doi: 10.1002/wics.1239

Ewing, C. J., Hausman, C. E., Pogacnik, J., Slot, J., and Bonello, P. (2019). Beech leaf disease: an emerging forest epidemic. *For. Pathol.* 49:e12488. doi: 10.1111/efp.12488

Ewing, C. J., Slot, J., Benítez, M.-S., Rosa, C., Malacrinò, A., Bennett, A., et al. (2021). The foliar microbiome suggests that fungal and bacterial agents may be involved in the beech leaf disease pathosystem. *Phytobiomes J.* 5, 335–349. doi: 10.1094/PBIOMES-12-20-0088-R

Fearer, C. J. (2022). *Investigating the Fagus Grandifolia-Beech Leaf Disease Pathosystem Using Metabarcoding, Phenological Observations, and Near-Infrared Spectroscopy*. Columbus, OH, USA: The Ohio State University.

Fearer, C. J., Conrad, A. O., Marra, R. E., Georskey, C., Villari, C., Slot, J., et al. (2022a). A combined approach for early in-field detection of beech leaf disease using near-infrared spectroscopy and machine learning. *Front. For. Glob. Change* 5:934545. doi: 10.3389/ffgc.2022.934545

Fearer, C. J., Volk, D., Hausman, C. E., and Bonello, P. (2022b). Monitoring foliar symptom expression in beech leaf disease through time. *For. Pathol.* 52:e12725. doi: 10.1111/efp.12725

Febrero-Bande, M., and de la Fuente, M. O. (2012). Statistical computing in functional data analysis: the R package Fda.Usc. *J. Stat. Softw.* 51, 1–28. doi: 10.18637/jss.v051.i04

Fenn, K., Conlon, C., Jones, M., Quail, M. A., Holroyd, N. E., Parkhill, J., et al. (2006). Phylogenetic relationships of the Wolbachia of nematodes and arthropods. *PLoS Pathog.* 2:e94. doi: 10.1371/journal.ppat.0020094

Figueiredo, G., Osório, H., Mendes, M. V., and Mendo, S. (2025). A review on the expanding biotechnological frontier of *Pedobacter*. *Biotechnol. Adv.* 82:108588. doi: 10.1016/j.biotechadv.2025.108588

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01

Galvan, F. E. R., Pavlick, R., Trolley, G., Aggarwal, S., Sousa, D., Starr, C., et al. (2023). Scalable early detection of grapevine viral infection with airborne imaging spectroscopy. *Phytopathology* 113, 1439–1446. doi: 10.1094/PHYTO-01-23-0030-R

Gold, K. M., Townsend, P. A., Chlus, A., Herrmann, I., Couture, J. J., Larson, E. R., et al. (2020). Hyperspectral measurements enable pre-symptomatic detection and differentiation of contrasting physiological effects of late blight and early blight in potato. *Remote Sens.* 12:286. doi: 10.3390/rs12020286

Gong, J.-T., Li, T.-P., Wang, M.-K., and Hong, X.-Y. (2023). Wolbachia-based strategies for control of agricultural pests. *Curr. Opin. Insect Sci.* 57:101039. doi: 10.1016/j.cois.2023.101039

Goraya, M., Kantor, C., Vieira, P., Martin, D., and Kantor, M. (2024). Deciphering the vectors: unveiling the local dispersal of Litylenchus Crenatae Ssp. Mccannii in the American beech (*Fagus Grandifolia*) Forest ecosystem. *PLoS One* 19:e0311830. doi: 10.1371/journal.pone.0311830

Grundler, F. M. W., and Hofmann, J. (2011). "Water and nutrient transport in nematode feeding sites," in *Genomics and Molecular Genetics of Plant-Nematode Interactions*, eds. J. Jones, G. Gheysen and C. Fenoll (Dordrecht, The Netherlands: Springer).

Guo, Q., Potter, K. M., Ren, H., and Zhang, P. (2023). Impacts of exotic pests on forest ecosystems: an update. *Forests* 14:605. doi: 10.3390/f14030605

Hardoim, P. R., Guerra, R., Rosa Da Costa, A. M., Serrano, M. S., Sánchez, M. E., and Coelho, A. C. (2016). Temporal metabolic profiling of the *Quercussuber – Phytophthoracinnamomi* system by middle-infrared spectroscopy. *Forest Pathol.* 46, 122–133. doi: 10.1111/efp.12229

Hassani, M. A., Durán, P., and Hacquard, S. (2018). Microbial interactions within the plant holobiont. *Microbiome* 6:58. doi: 10.1186/s40168-018-0445-0

Ho, C.-S., Jean, N., Hogan, C. A., Blackmon, L., Jeffrey, S. S., Holodniy, M., et al. (2019). Rapid identification of pathogenic Bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* 10:4927. doi: 10.1038/s41467-019-12898-9

Huang, Q., Yang, M., Ouyang, L., Wang, Z., and Lin, J. (2025). Vis/NIR spectroscopy and chemometrics for non-destructive estimation of chlorophyll content in different plant leaves. *Sensors* 25:1673. doi: 10.3390/s25061673

Johnston, K. L., and Taylor, M. J. (2007). *Wolbachia* in filarial parasites: targets for filarial infection and disease control. *Curr. Infect. Dis. Rep.* 9, 55–59. doi: 10.1007/s11908-007-0023-2

Kantor, C., Demirel, M. C., and Kantor, M. (2025). Unveiling the threat of beech leaf disease: lessons from North America. *Front. For. Glob. Change* 8:1606260. doi: 10.3389/ffgc.2025.1606260

Kanzaki, N., Ichihara, Y., Aikawa, T., Ekino, T., and Masuya, H. (2019). *Litylenchuscrenatae* n. sp. (Tylenchomorpha: Anguinidae), a leaf gall nematode parasitising *Faguscrenata* blume. *Nematology* 21, 5–22. doi: 10.1163/15685411-00003190

Kaplan, F., Badri, D. V., Zachariah, C., Ajredini, R., Sandoval, F. J., Roje, S., et al. (2009). Bacterial attraction and quorum sensing inhibition in *Caenorhabditis Elegans* exudates. *J. Chem. Ecol.* 35, 878–892. doi: 10.1007/s10886-009-9670-0

Korenblum, E., and Aharoni, A. (2019). Phytobiome metabolism: beneficial soil microbes steer crop plants' secondary metabolism. *Pest Manag. Sci.* 75, 2378–2384. doi: 10.1002/ps.5440

Kunda, P., Mondal, S., De, D., Dhal, P. K., and Mukherjee, A. (2024). Alteration of Rice root endophytic bacterial community composition by Meloidogyne Graminicola and identification of potential biocontrol agent. *Ann. Microbiol.* 74:42. doi: 10.1186/s13213-024-01789-0

Kusonmano, K., Netzer, M., Baumgartner, C., Dehmer, M., Liedl, K. R., and Graber, A. (2012). Effects of pooling samples on the performance of classification algorithms: a comparative study. *Sci. World J.* 2012:278352.

Leontopoulos, S. V., Petrotos, K., Anatolioti, V., Skenderidis, P., Tsilfoglou, S., and Vagelas, I. (2017). Chemotactic responses of Pseudomonas Oryzihabitans and second stage juveniles of Meloidogyne Javanica on tomato root tip exudates. *Int. J. Food Biosyst. Eng.* 5, 75–100.

Li, Z., Yang, X., Zhang, D., Shi, X., Lei, L., Zhou, F., et al. (2025). Exploration of Oral microbiota alteration and AI-driven non-invasive hyperspectral imaging for CAD prediction. *BMC Cardiovasc. Disord.* 25:102. doi: 10.1186/s12872-025-04555-5

Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. Vienna, Austria: R Foundation for Statistical Computing. vol. 2.

Liland, K. H., Mevik, B.-H., and Wehrens, R. (2025). PLS: Partial Least Squares and Principal Component Regression. Available online at: https://github.com/khliland/pls (Accessed December 18, 2025).

Liu, Z., Parida, S., Prasad, R., Pandey, R., Sharma, D., and Barman, I. (2022). Vibrational spectroscopy for decoding Cancer microbiota interactions: current evidence and future perspective. *Semin. Cancer Biol.* 86, 743–752. doi: 10.1016/j.semcancer.2021.07.004

Lu, X., and Rasco, B. A. (2012). Determination of antioxidant content and antioxidant activity in foods using infrared spectroscopy and chemometrics: a review. *Crit. Rev. Food Sci. Nutr.* 52, 853–875. doi: 10.1080/10408398.2010.511322

Lyu, X., Li, P., Jin, L., Yang, F., Pucker, B., Wang, C., et al. (2024). Tracing the evolutionary and genetic footprints of atmospheric Tillandsioids transition from land to air. *Nat. Commun.* 15:9599. doi: 10.1038/s41467-024-53756-7

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., et al. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* 17:e1009442. doi: 10.1371/journal.pcbi.1009442

Mannaa, M., Park, I., and Seo, Y.-S. (2018). Genomic features and insights into the taxonomy, virulence, and benevolence of plant-associated Burkholderia species. *Int. J. Mol. Sci.* 20:121. doi: 10.3390/ijms20010121

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Masaitis, G., Mozgeris, G., and Augustaitis, A. (2013). Spectral reflectance properties of healthy and stressed coniferous trees. *iForest Biogeosci. For.* 6, 30–36. doi: 10.3832/ifor0709-006

McIntire, C. D. (2023). Physiological impacts of beech leaf disease across a gradient of symptom severity among understory American beech. *Front. Forests Global Change* 6:1146742. doi: 10.3389/ffgc.2023.1146742

Meyer, D., Dimitriadou, E., Hornik, K., et al (2019). "E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien." R Package Version 1 (2). ed. T. U. Wien (Vienna, Austria: Vienna University of Technology).

Miles, A. (2023). "*From Forested to Urban Ecosystems: Refining Tree Health in American Elm (Ulmus Americana) and American Beech (Fagus Grandifolia)*." *M.S. Thesis*. University Park, PA, USA: The Pennsylvania State University.

Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., et al. (2019). The UNITE database for molecular identification of Fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47, D259–D264. doi: 10.1093/nar/gky1022

Oksanen, J., Simpson, G. L., and Blanchet, F. G. (2025). Vegan: Community Ecology Package. Available online at: https://vegandevs.github.io/vegan/

Paap, T., Wingfield, M. J., Burgess, T. I., Hulbert, J. M., and Santini, A. (2020). Harmonising the fields of invasion science and forest pathology. *NeoBiota* 62, 301–332. doi: 10.3897/neobiota.62.52991

Pang, Z., Chen, J., Wang, T., Gao, C., Li, Z., Guo, L., et al. (2021). Linking plant secondary metabolites and plant microbiomes: a review. *Front. Plant Sci.* 12:621276. doi: 10.3389/fpls.2021.621276

Petrikovszki, R., Cseresnyés, I., Bárányos, F., Molnár, A., and Boros, G. (2023). Early detection of root-knot nematode (*Meloidogyne incognita*) infection by monitoring root dielectric response non-destructively. *Int. Agrophys.* 37, 179–184. doi: 10.31545/intagr/162798

Pramreiter, M., and Grabner, M. (2023). The utilization of European beech wood (*Fagus sylvatica* L.) in Europe. *Forests* 14:1419. doi: 10.3390/f14071419

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Razaque, A., Frej, M. B. H., Almi'ani, M., Alotaibi, M., and Alotaibi, B. (2021). Improved support vector machine enabled radial basis function and linear variants for remote sensing image classification. *Sensors* 21:4431. doi: 10.3390/s21134431

Reed, S. E., Greifenhagen, S., Yu, Q., Hoke, A., Burke, D. J., Carta, L. K., et al. (2020). Foliar nematode, *Litylenchus crenatae* ssp. *mccannii*, population dynamics in leaves and buds of beech leaf disease-affected trees in Canada and the US. *For. Pathol.* 50:e12599. doi: 10.1111/efp.12599

Rohart, F., Gautier, B., Singh, A., and Cao, K.-A. L. (2017). mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13:e1005752. doi: 10.1371/journal.pcbi.1005752

Ross, M. (2013). Know your Deer Plants: American Beech. National Deer Association. Available online at: https://deerassociation.com/know-deer-plants-american-beech/

Roussin-Léveillée, C., Mackey, D., Ekanayake, G., Gohmann, R., and Moffett, P. (2024). Extracellular niche establishment by plant pathogens. *Nat. Rev. Microbiol.* 22, 360–372. doi: 10.1038/s41579-023-00999-8

Sánchez-Moreno, S., and Ferris, H. (2025). Nematodes as Bioindicators. Oxfordshire, England: CABI.

Selvi, E., Liu, D., and Bonello, P. (2025). Anticipating shifts in American beech distribution in a changing climate. *J. Biogeogr.* 52:e70010. doi: 10.1111/jbi.70010

Šigutová, H., Šigut, M., Pyszko, P., Kostovčík, M., Kolařík, M., and Drozd, P. (2023). Seasonal shifts in bacterial and fungal microbiomes of leaves and associated Leaf-Mining larvae reveal persistence of Core taxa regardless of diet. *Microbiol. Spectrum* 11, e0316022–e0316022. doi: 10.1128/spectrum.03160-22

Silby, M. W., Winstanley, C., Godfrey, S. A. C., Levy, S. B., and Jackson, R. W. (2011). *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol. Rev.* 35, 652–680. doi: 10.1111/j.1574-6976.2011.00269.x

Slatko, B. E., Taylor, M. J., and Foster, J. M. (2010). The Wolbachia endosymbiont as an anti-filarial nematode target. *Symbiosis* 51, 55–65. doi: 10.1007/s13199-010-0067-1

Stephanson, C. A., and Coe, N. R. (2017). Impacts of beech bark disease and climate change on American beech. *Forests* 8:155. doi: 10.3390/f8050155

Stevens, A., and Ramirez-Lopez, L. (2025). An Introduction to the Prospectr Package

Straub, D., Blackwell, N., Langarica-Fuentes, A., Peltzer, A., Nahnsen, S., and Kleindienst, S. (2020). Interpretations of environmental microbial community studies are biased by the selected 16S rRNA (gene) amplicon sequencing pipeline. *Front. Microbiol.* 11:550420. doi: 10.3389/fmicb.2020.550420

Thomas, B. O., Lechner, S. L., Ross, H. C., Joris, B. R., Glick, B. R., and Stegelmeier, A. A. (2024). Friends and foes: Bacteria of the hydroponic plant microbiome. *Plants* 13:3069. doi: 10.3390/plants13213069

Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F. F., et al. (2020). *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nat. Rev. Microbiol.* 18, 415–427. doi: 10.1038/s41579-020-0361-8

Türker-Kaya, S., and Huck, C. (2017). A review of mid-infrared and near-infrared imaging: principles, concepts and applications in plant tissue analysis. *Molecules* 22:168. doi: 10.3390/molecules22010168

Viana, L. R., Silveira, F. A. O., Santos, J. C., Rosa, L. U. I. Z. H. E. N. R. I. Q. U. E., Cares, J. U. V. E. N. I. L. E., Café-Filho, A. D. A. L. B. E. R. T. O. C. O. R. R. Ê. A., et al. (2013). Nematode-induced galls in *Miconia albicans*: effect of host plant density and correlations with performance. *Plant Species Biol.* 28, 63–69. doi: 10.1111/j.1442-1984.2011.00358.x

Vieira, P., Kantor, M. R., Jansen, A., Handoo, Z. A., and Eisenback, J. D. (2023). Cellular insights of beech leaf disease reveal abnormal ectopic cell division of symptomatic interveinal leaf areas. Preprint. *Cell Biol.* doi: 10.1101/2023.06.22.546113

Wasala, S. K., Brown, A. M. V., Kang, J., Howe, D. K., Peetz, A. B., Zasada, I. A., et al. (2019). Variable abundance and distribution of Wolbachia and Cardinium endosymbionts in plant-parasitic nematode field populations. *Front. Microbiol.* 10:964. doi: 10.3389/fmicb.2019.00964

Wasala, S. K., Hesse, C., Wram, C. L., Howe, D. K., Zasada, I. A., and Denver, D. R. (2023). Unraveling microbial endosymbiosis dynamics in plant-parasitic nematodes with a genome skimming strategy. *Appl. Microbiol.* 3, 1229–1248. doi: 10.3390/applmicrobiol3040085

Weyer, L. G., and Lo, S.-C. (2002). "Spectra-structure correlations in the near-infrared," in Handbook of Vibrational Spectroscopy, ed. J. Chalmers (Chichester, West Sussex, United Kingdom: John Wiley & Sons, Ltd), 3, 1817–1837.

Williams, G. M., Ginzel, M. D., Ma, Z., et al. (2023). The global forest health crisis: a public-good social dilemma in need of international collective action. *Annu. Rev. Phytopathol.* 61, 377–401.

Workman, J. Jr., and Weyer, L. (2007). *Practical Guide to Interpretive Near-Infrared Spectroscopy*. Boca Raton: CRC Press.

Xu, S., Zhan, L., Tang, W., Wang, Q., Dai, Z., Zhou, L., et al. (2023). MicrobiotaProcess: a comprehensive R package for deep mining microbiome. *Innovation* 4:100388. doi: 10.1016/j.xinn.2023.100388

Yergaliyev, T. M., Alexander-Shani, R., Dimerets, H., Pivonia, S., Bird, D. M. K., Rachmilevitch, S., et al. (2020). Bacterial community structure dynamics in *Meloidogyne incognita* -infected roots and its role in worm-microbiome interactions. *mSphere* 5. doi: 10.1128/msphere.00306-20

Yu, R., Huo, L., Huang, H., Yuan, Y., Gao, B., Liu, Y., et al. (2022). Early detection of pine wilt disease tree candidates using time-series of spectral signatures. *Front. Plant Sci.* 13:1000093. doi: 10.3389/fpls.2022.1000093

Zaneveld, J. R., McMinds, R., and Thurber, R. V. (2017). Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nat. Microbiol.* 2:17121. doi: 10.1038/nmicrobiol.2017.121

Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., da Rocha, U. N., Shi, S., et al. (2018). Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* 3, 470–480. doi: 10.1038/s41564-018-0129-3

Zhao, D., You, Y., Fan, H., Zhu, X., Wang, Y., Duan, Y., et al. (2018). The role of sugar transporter genes during early infection by root-knot nematodes. *Int. J. Mol. Sci.* 19:302. doi: 10.3390/ijms19010302

Zou, D., Zhang, C., Liu, Y., and Li, M. (2024). Biogeographical distribution and community assembly of Myxococcota in mangrove sediments. *Environ. Microbiome* 19:47. doi: 10.1186/s40793-024-00593-2