Microbial biomarkers of tree water status for next-generation biomonitoring of forest ecosystems

Marine C Cambon, Marine Trillat, Isabelle Lesur-kupin, Régis Burlett, Emilie Chancerel, Erwan Guichoux, Lucie Piouceau, Bastien Castagneyrol, Grégoire Le Provost, Stéphane Robin, et al.

Received: 8 March 2023 | Revised: 14 September 2023 | Accepted: 19 September 2023

DOI: 10.1111/mec.17149

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Microbial biomarkers of tree water status for next-generation biomonitoring of forest ecosystems

Marine C. Cambon[1,2] | Marine Trillat[1] | Isabelle Lesur-Kupin[1,3] | Régis Burlett[1] | Emilie Chancerel[1] | Erwan Guichoux[1] | Lucie Piouceau[1] | Bastien Castagneyrol[1] | Grégoire Le Provost[1] | Stéphane Robin[4] | Yves Ritter[1] | Inge Van Halder[1] | Sylvain Delzon[1] | David A. Bohan[5] | Corinne Vacher[1]

[1]INRAE, University of Bordeaux, BIOGECO, Pessac, France

[2]School of Natural Sciences, Bangor University, Bangor, UK

[3]HelixVenture, Mérignac, France

[4]CNRS, LPSM, Sorbonne Université, Paris, France

[5]Agroécologie, INRAE, Université Bourgogne Franche-Comté, Dijon, France

Correspondence
Corinne Vacher, INRAE, University of Bordeaux, BIOGECO, Pessac, France.
Email: corinne.vacher@inrae.fr

Marine Cambon, Bangor University, Bangor, UK.
Email: cambonmarine@gmail.com

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR-17-CE32-0011, ANR-16-CE32-0003-01 and ANR-10-LABX-25-01; Conseil Régional Aquitaine, Grant/Award Number: 2016-1R20301-00007218

Handling Editor: Pierre Taberlet

Abstract

Next-generation biomonitoring proposes to combine machine-learning algorithms with environmental DNA data to automate the monitoring of the Earth's major ecosystems. In the present study, we searched for molecular biomarkers of tree water status to develop next-generation biomonitoring of forest ecosystems. Because phyllosphere microbial communities respond to both tree physiology and climate change, we investigated whether environmental DNA data from tree phyllosphere could be used as molecular biomarkers of tree water status in forest ecosystems. Using an amplicon sequencing approach, we analysed phyllosphere microbial communities of four tree species (*Quercus ilex*, *Quercus robur*, *Pinus pinaster* and *Betula pendula*) in a forest experiment composed of irrigated and non-irrigated plots. We used these microbial community data to train a machine-learning algorithm (Random Forest) to classify irrigated and non-irrigated trees. The Random Forest algorithm detected tree water status from phyllosphere microbial community composition with more than 90% accuracy for oak species, and more than 75% for pine and birch. Phyllosphere fungal communities were more informative than phyllosphere bacterial communities in all tree species. Seven fungal amplicon sequence variants were identified as candidates for the development of molecular biomarkers of water status in oak trees. Altogether, our results show that microbial community data from tree phyllosphere provides information on tree water status in forest ecosystems and could be included in next-generation biomonitoring programmes that would use in situ, real-time sequencing of environmental DNA to help monitor the health of European temperate forest ecosystems.

KEYWORDS
environmental DNA data, machine-learning algorithms, molecular biomarkers, next-generation biomonitoring, phyllosphere microbial communities, water stress

1 | INTRODUCTION

Next-generation biomonitoring of ecosystem change proposes to automate the monitoring of the Earth's major ecosystems by combining in situ, real-time sequencing of environmental DNA with machine-learning algorithms (Bohan et al., 2017; Cordier et al., 2021; Cordier, Lanzén, et al., 2018). As traditional biomonitoring, next-generation biomonitoring relies on biomarkers, defined as organisms or communities of organisms whose reactions give clues for the condition of the whole ecosystem (Gerhardt, 2002). However, next-generation biomonitoring proposes to derive these biomarkers from environmental DNA data (Cordier et al., 2017; Cordier, Forster, et al., 2018; Cordier, Lanzén, et al., 2018), rather than from species inventories based on morphological observations. To our knowledge, the concept of next-generation biomonitoring has hardly been applied to forest ecosystems so far, although forest ecosystems are increasingly impacted by human activity and climate change. Drought stress is a major driver of forest ecosystem change and is responsible for tree mortality events all around the world (Allen et al., 2010, 2015; Brodribb et al., 2020). Predicting the occurrence of these tree mortality events requires global, interdisciplinary, real-time and long-term monitoring approaches, integrating multiple indicators of tree drought stress (Hartmann et al., 2018). Here we therefore searched for novel molecular biomarkers of tree water status, derived from environmental DNA, that could be included in integrative programmes of European temperate forest ecosystem monitoring.

Our vision of the implementation of next-generation biomonitoring for forest ecosystems relies on the diversified communities of microorganisms that constitute the microbiota of all plant and tree species. The plant microbiota includes some microbial species that promote plant growth (Compant et al., 2010; Hardoim et al., 2015), and contribute to plant resistance to microbial pathogens (Hacquard et al., 2017; Hacquard & Schadt, 2015; McLaren & Callahan, 2020; Vannier et al., 2019) and insect pests (Pineda et al., 2017; Vacher et al., 2021), and to the tolerance to abiotic stressors including drought (Lata et al., 2018; Rho et al., 2018; Rodriguez et al., 2008). In addition to contributing to plant response to environmental stressors, plant-associated microbial communities respond rapidly to environmental change. For instance, it has been shown that root bacterial communities of trees respond to drought by a richness decrease (Kristy et al., 2022). We therefore hypothesised that the tree microbiota could be a relevant biomarker of tree condition in forest ecosystems, as it responds rapidly to environmental change while contributing to tree health and growth.

Among all the microbial communities associated with a tree, those of the leaf are of particular interest for next-generation biomonitoring of forest ecosystems for two reasons. First, leaves are an easy-to-sample above-ground material. Moreover, epiphytes (i.e. the microbes living on the leaf surface) are at the interface between the plant and the atmosphere, while endophytes (i.e. the microbes living within leaf tissue) are closely linked to the leaf condition (Vacher et al., 2016). Therefore, the phyllosphere microbiota (i.e. the total community of endophytes and epiphytes) responds to variations in both climate (Aydogan et al., 2018; Cordier et al., 2012; Laforest-Lapointe et al., 2017) and plant physiology (Kembel et al., 2014; Rosado et al., 2018). The phyllosphere microbiota of trees is thus expected to be impacted by climate change, especially drought events and temperature rises (Perreault & Laforest-Lapointe, 2022; Zhu et al., 2022). Accordingly, several experimental studies in forest ecosystems have demonstrated that the phyllosphere microbiota responds significantly to drought events and climate warming. For instance, the diversity of phyllosphere bacterial and fungal communities in holm oak, *Quercus ilex*, has been found to increase with drought stress (Peñuelas et al., 2012). Peñuelas et al. (2012) suggested that the increase in volatile organic compound emissions after a moderate drought might raise the amount and diversity of carbon sources available to microorganisms at the leaf surface, thus leading to an increase in diversity of the total microbial community. In contrast, leaf fungal diversity has been shown to decrease with heat in pedunculate oak, *Quercus robur* (Faticov et al., 2021) and poplar, *Populus balsamifera* (Bálint et al., 2015).

The analysis of microbial communities in environmental samples, such as tree leaves, has been greatly facilitated by high-throughput sequencing methods developed over the last 15 years. The assessment of microbial community composition is increasingly cheaper and faster (Nilsson et al., 2019). As environmental sequencing data are big data that hold a huge amount of information, they have required the development of advanced computational methods to extract relevant information. For instance, machine-learning algorithms are increasingly used to detect changes in environmental conditions or host physiology, based on the microbial community composition in environmental DNA samples (see Knights et al., 2011; Namkung, 2020 for a review). One of the most widely used machine-learning methods is the Random Forest algorithm (Breiman, 2001; Xu et al., 2022), which is a supervised classification algorithm requiring two steps. First, a small microbial community dataset (i.e. the training dataset) is used to train the Random Forest algorithm to classify samples into groups, defined by a discrete variable characterizing the environment where the samples were collected. Second, the algorithm is used to predict the variable of interest for other samples, for which only the microbial community composition is known. For instance, Random Forest algorithms have been used on microbial community data to classify sites according to their pollution level (Liu et al., 2018) or the potential for crop productivity (Chang et al., 2017), and to assign seeds to a plant variety (Kim et al., 2020). In addition to the ability to handle huge datasets, Random Forest algorithms allow the construction of decision-making tools for ecosystem monitoring (Cordier et al., 2017, 2021).

The aim of this study was to investigate whether microbial environmental DNA, sequenced with high-throughput methods and analysed with machine-learning algorithms, can be used as a biomarker of tree water status, in order to develop next-generation biomonitoring programmes for forest ecosystems. We focused on phyllosphere microbiota and analysed its association with tree water status in four tree species (*Q. ilex*, *Q. robur*, *Pinus pinaster* and *Betula pendula*). We investigated which component of microbial community data performs best at detecting tree water status. We specifically compared the ability of fungal versus bacterial data and rare versus abundant taxa at classifying trees according to their water status. We also investigated which taxonomic aggregation levels (ASV, genus, family, class, order) of the microbial community data perform best. We finally discussed the strengths and limitations of these DNA-based biomarkers.

2 | MATERIALS AND METHODS

2.1 | Study site

The study was carried out in the ORPHEE (https://sites.google.com/view/orpheeexperiment/home) tree diversity experiment. This experiment, planted in 2008 in the south-west of France (44°44′24.9″N, 0°47′48.1″W), belongs to the TreeDivNet international network (https://treedivnet.ugent.be/). Eight block repetitions were established, each block consisting of a set of 32 experimental forest plots planted with a combination of one to five temperate tree species (*P. pinaster*, *B. pendula*, *Q. robur*, *Q. ilex* and *Q. pyrenaica*). Each plot is 0.4 ha and contains 100 trees planted 2 m apart in a 10 × 10 grid. The whole experiment covers 12 ha. Four out of eight blocks have been irrigated since 2015 using a 2 m high sprinkling system (Figure 1a). Irrigated blocks receive 3 mm of water every night from the 22th of May to late September. This volume was estimated based on evapotranspiration data and has been proved sufficient to avoid irrigated trees suffering from soil water deficit during the growing season in several studies conducted before and after our sampling. In 2015, *Q. robur* pre-dawn water potential was lower in non-irrigated blocks than in irrigated blocks (−0.37 and −0.19 MPa respectively; Castagneyrol et al., 2017). The same pattern was observed for *B. pendula* in 2016 (−1.63 and −0.43 MPa for non-irrigated and irrigated blocks respectively; Castagneyrol et al., 2018) and for *Q. ilex* in 2019 (−0.49 and −0.18 MPa for non-irrigated and irrigated blocks respectively; Galmán et al., 2022).

2.2 | Leaf sampling and processing

We analysed the leaf microbiota of two evergreen tree species (*P. pinaster* and *Q. ilex*) and two deciduous tree species (*B. pendula* and *Q. robur*) planted in monocultures in the ORPHEE experiment. The trees belonged to 32 plots, corresponding to 4 tree species × 2 water treatments (irrigated vs. non-irrigated) × 4 replicates (blocks). In each plot, we sampled leaves (or needles) on 12 trees selected in the central part of the plot to limit edge effects (Figure 1a).

We performed two sampling campaigns. The first one occurred in May 2019, a few days after irrigation started, and the second in

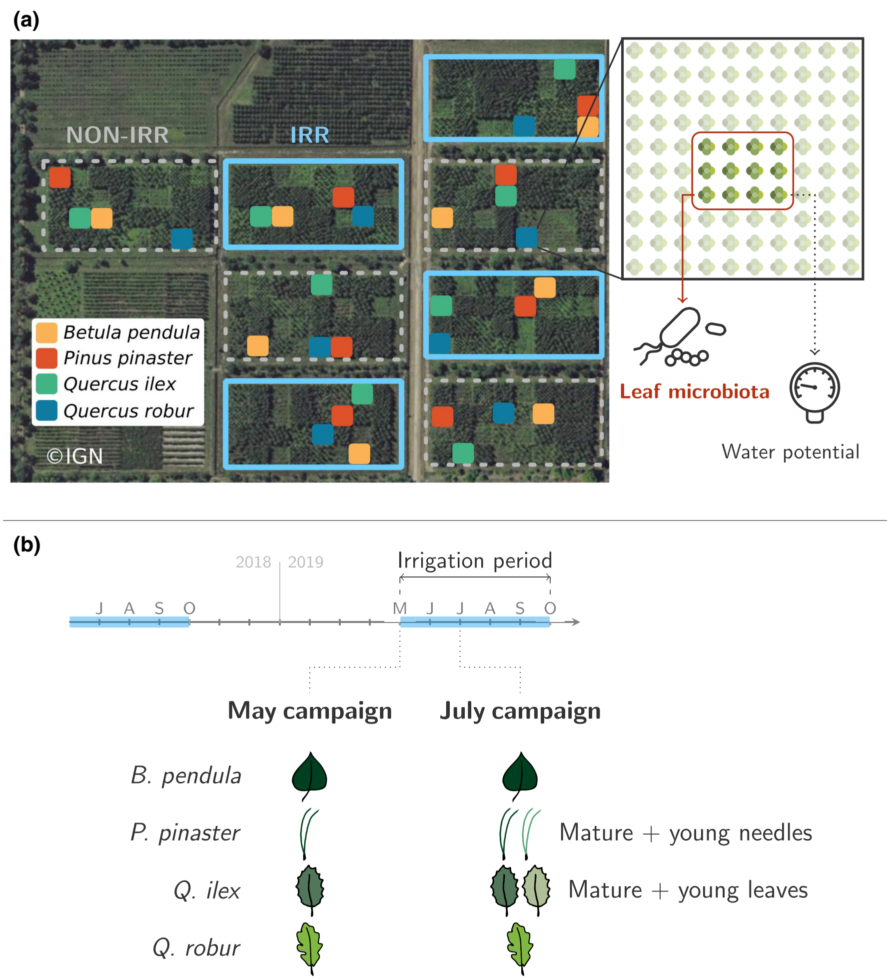

FIGURE 1 Experimental design. (a) The ORPHEE experimental setup. For each tree species (*Betula pendula*, *Pinus pinaster*, *Quercus ilex* and *Quercus robur*), four irrigated (IRR) and four non-irrigated (NON-IRR) monoculture plots were sampled. In each plot, leaves from 12 trees were sampled in May and July 2019 for microbiota analysis. Branches from three trees were sampled to measure predawn water potential. (b) Timeline of the sampling. Irrigated plots were watered between early May and late September (blue line) in the sampling year (2019) and during the 4 years preceding sampling (2015 to 2018). Leaves were sampled in May, before the summer period, and in July, during the summer. *Q. robur* and *B. pendula* being deciduous species, leaves sampled in both May and July were leaves of the year. *P. pinaster* and *Q. ilex* being evergreen species, leaves/needles sampled in May were mature leaves/needles from the previous spring, while leaves sampled in July were both mature leaves/needles (dark green) and young leaves/needles (light green).

July 2019 (Figure 1b). In May, the sampling campaign lasted two consecutive days (23th and 24th of May 2019), all the trees of a given species being sampled the same day. The July sampling lasted 3 days (16th–18th of July 2019), with the sampling of each species completed in 2 days.

For each sampling campaign, a south-facing branch was cut from the canopy of each tree, at approximately 1.5 m height for *Q. ilex* and *Q. robur* and 8 m height for *B. pendula* and *P. pinaster*, using shears and pole pruners respectively. During the May sampling campaign, three visually healthy leaves (or needle pairs) were sampled from each branch, using new gloves between each tree. In July, three additional leaves were sampled in *Q. ilex*, to obtain thee young and three mature leaves per tree. Similarly, three additional needle pairs were sampled in *P. pinaster* (Figure 1b).

Leaves and needles were processed in the field, immediately after sampling. One disc of 5 or 7 mm diameter was cut in each leaf for *B. pendula* and the two oak species, respectively, while eight chunks of approximately 2 mm each were cut from each needle pair in *P. pinaster*. Tools were cleaned with 10% bleach and 70% ethanol between each tree and an autoclaved piece of paper filter was used as a work surface. The three leaf discs or 24 needle chunks from the same tree were placed together into 1 mL of RNA*later* (Invitrogen) in a 2 mL Eppendorf tube and stored on ice to prevent DNA degradation. The same day, samples were frozen at −80°C.

In addition, we analysed the endophytic and epiphytic communities of leaves for three trees per plot, chosen randomly in the centre of the plot among the 12 trees sampled previously (Figure 1a). To sample the epiphytic community, the upper and lower surfaces of three leaves per tree, or three needle pairs per tree, were swabbed in the field using sterile swabs previously soaked in sterile RNA*later*. Swabs were then stored on ice until storage at −20°C in the lab. To sample the endophytic community, three leaves or three needle pairs per tree were collected and placed into plastic bags. The bags were stored on ice in the field before storage at −80°C in the lab a couple hours later. Leaves and needles were subsequently surface-sterilized, with a slightly modified version of Unterseher and Schnittler (2009)'s protocol. Leaves and needles were (i) washed with sterile distilled water, then placed in (ii) 70% EtOH for 2 min, (iii) 2% $Ca(ClO)_2$ for 5 min, (iv) 70% EtOH for 1 min and (v) briefly rinsed in commercial sterile purified water (Otec Aguettant, France). Leaf discs and needle chunks were then cut as described above and stored at −20°C until DNA extraction.

2.3 | Water potential measurements

Predawn water potential was measured on three trees randomly chosen in the centre of each plot (Figure 1a), both during the May and July sampling campaigns (Figure 1b). The three trees were the same as those selected for the analysis of epiphytic and endophytic microbial communities. Water potential measurements and leaf sampling were performed on the same day. Water potential measurements were performed between 5:00 AM and 7:00 AM by installing two Scholander pressure bombs (DGMeca, Gradignan, France) in the centre of the ORPHEE experiment. One branch was collected from the south side of the upper crown of each tree and water potentials were measured within 1 min of branch collection. The value of water potential was estimated as the negative of the balance pressure (MPa) applied on leaves using pressure chambers.

2.4 | Leaf microbial community profiling

2.4.1 | DNA extraction

Leaf discs and needle chunks were taken out from the storage solution (RNA*later*, Invitrogen) under a laminar flow hood using sterile tools. The excess solution was removed using an autoclaved piece of paper filter. For each tree, a sample consisted of either three leaf discs or 24 needle chunks. For each tree species, samples were randomized in 96-well plates and stored at −80°C. Leaf samples were then frozen in liquid nitrogen and ground for 3 × 30′ at 1500 rpm using two 4 mm diameter autoclaved steel beads per well with a Geno Grinder (SPEX Group Holdings Ltd, Aberdeen, UK). Needle samples were freeze-dried overnight prior to grinding to facilitate sample disruption. DNA was extracted with a DNeasy Plant Mini Kit 96 (Qiagen) following the manufacturer's protocol, except that the incubation time was extended to 1 h at 65°C. All extractions were made in a confined laboratory to prevent any contamination.

Epiphyte swab tips were cut using scissors under a laminar flow hood into 2 mm pieces and stored in 2 mL Eppendorf tubes at −80°C until DNA extraction. Scissors were cleaned with 10% bleach and 70% ethanol between each sample. DNA was extracted using the PowerSoil kit (Qiagen) following the manufacturer's protocol.

Three negative extraction controls were placed on each extraction plate. They consisted of wells without any plant material, containing only the extraction reagents. DNA yield and purity were checked by spectrophotometry using Nanodrop 2000 (ThermoScientific), and electrophoresis on 2% agarose gels. DNA yields obtained for *B. pendula* and *P. pinaster* were lower than the ones obtained for *Q. robur* and *Q. petrae*.

2.4.2 | DNA amplification

The V5-V6 hypervariable region of the bacterial 16S rRNA gene and the ITS1 region of the fungal nuclear ribosomal internal transcribed spacer (ITS) were then amplified from all samples. All amplifications were made in a confined laboratory to prevent any contamination.

The V5-V6 region of the bacterial 16S rRNA gene was amplified using the 799F (Chelius & Triplett, 2001) and 1115R

(Turner et al., 1999) primers. Each primer contained the Illumina adaptor sequence, a tag and a heterogeneity spacer, as described in (Laforest-Lapointe et al., 2017) (799F: 5′-CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxxxxxxxxxxHS-AACMGGATTAGATACCCKG-3′; 1115R: 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxxxxxxxxxHS-AGGGTTGCGCTCGTTG-3′, where HS represents a 0-7-base-pair heterogeneity spacer and "x" a 12 nucleotides tag). For *Q. ilex* and *Q. robur*, the reactions were performed in a volume of 20 μL, containing 10 μL of the 2X Taq polymerase (Qiagen), 1 μL of each primer (0,1 μM final) and 1 μL of environmental DNA. DNA was first denatured at 95°C for 15 min, and then amplified for 30 cycles of 94°C for 30 s, 53°C for 90 s, 72°C for 90 s, and finally extended at 72°C for 10 min. For *B. pendula* and *P. pinaster*, the reactions were performed in 20 μL, containing 4 μL of Buffer Phusion HF 5X, 0.2 μL of Phusion HSII polymerase (New England BioLab), 0.6 μL of dimethyl sulfoxide (Thermo Scientific), 2 μL of each primer (0,2 μM final) and 1 μL of environmental DNA. DNA was first denatured at 98°C for 30 s, and then amplified for 35 cycles of 98°C for 15 s, 60°C for 30 s, 72°C for 30 s and finally extended at 72°C for 10 min.

The fungal ITS1 region was amplified using the ITS1F (Gardes & Bruns, 1993) and ITS2 (White et al., 1990) primers containing the Illumina adaptor sequence and a tag (ITS1F: 5′-CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxxxxxxxxxxCTTGGTCATTTAGAGGAAGTAA-3′; ITS2: 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxxxxxxxxxGCTGCGTTCTTCATCGATGC-3′, where "x" is a 12 nucleotides tag). For all tree species, the reactions were performed in a volume of 20 μL, containing 10 μL of the 2X Taq polymerase (Qiagen), 1 μL of each primer (0,1 μM final) and 1 μL of environmental DNA. DNA was first denatured at 95°C for 15 min, and then amplified for 35 cycles of 94°C for 30 s, 57°C for 90 s, 72°C for 90 s, and finally extended at 72°C for 10 min.

For all PCR reactions, negative PCR controls (3 per plate) consisted in wells without any DNA, containing only the PCR reagents. Positive PCR controls (1 per plate) consisted in pure DNA of a bacterial or a fungal marine species (*Sulfitobacter pontiacus* and *Candida oceani* respectively), as they were unlikely to be found in the tree leaf samples. PCR was conducted on a Veriti 96-well Thermal Cycler (Applied Biosystems) and all amplifications were confirmed by electrophoresis on a 2% agarose gel.

PCR products were purified using SpeedBeads™ magnetic carboxylate modified particles (Sigma), quantified (Quant-it PicoGreen dsDNA assay kit, Thermo Fisher Scientific) and equimolarly pooled (Hamilton Microlab STAR robot). Amplicon concentrations obtained for *B. pendula* and *P. pinaster* were lower than the ones obtained for *Q. robur* and *Q. petrae*. Average size fragment was checked using a Tapestation (Agilent Technologies). Libraries were sequenced on a MiSeq (Illumina) with the reagent kit v2 (500-cycles, MS-102-2003) for leaf samples, and the nano reagent kit v2 (500-cycles, MS-103-1003) for swab samples.

2.4.3 | Bioinformatic analysis

Sequence demultiplexing (with exact index search) was performed using DoubleTagDemultiplexer (https://github.com/yoann-dufresne/DoubleTagDemultiplexer). Demultiplexed sequences were processed using the dada2 R package v1.12.1 (Callahan et al., 2016) following the dada2 tutorial (v1.16 and v1.8 for 16S and ITS respectively). Primer sequences were removed using cutadapt (Martin, 2011), and sequences with more than one expected error based on quality scores (Edgar & Flyvbjerg, 2015), containing ambiguous nucleotides or shorter than 100 bp were trimmed. Amplicon sequence variants (ASVs) were then inferred for each sample using the dada function. Forward and reverse denoised reads were paired using the mergePairs function, and chimeric sequences were removed using the removeBimeraDenovo function provided in the dada2 package. Taxonomic assignments of ASVs were performed with RDP Naive Bayesian Classifier algorithm (Wang et al., 2007) implemented in dada2 (assignTaxonomy function), with the SILVA reference database v138 (Quast et al., 2012) for bacteria and the UNITE reference database v8.3 (Nilsson et al., 2018; UNITE Community, 2019) for fungi. All ASVs unassigned to the bacterial kingdom or the Dikarya clade, or matching chloroplastic and mitochondrial sequences were removed. The ASV table was then curated using positive and negative controls. Negative controls were used to remove reads resulting from cross contamination or reagent contamination following the (Galan et al., 2016) procedure. Sequences detected in the positive controls were removed from other environmental samples following Galan et al. (2016) procedure. Additional contaminants were identified based on their frequency using the decontam package v1.12.0 (Davis et al., 2018). Finally, samples with less than 100 reads were discarded from the ASV table (Table S1).

2.5 | Statistical analysis

2.5.1 | Comparison of tree water status between irrigated and non-irrigated plots

Data analyses were performed using R version 4.0.2 (2020-06-22) (R Core Team, 2020). Predawn water potential measurements were analysed for each tree species with a linear mixed-effect model (LMM) using the lmer function of the lmerTest package v3.1.3 (Kuznetsova et al., 2017), followed by an analysis of variance (ANOVA), and a *post hoc* Wilcoxon rank-sum test with continuity correction using the wilcox.test function (R base). The model had the sampling month, the irrigation treatment and their interaction as fixed effects, and the block as a random factor.

2.5.2 | Comparison of phyllosphere microbiota diversity and composition between irrigated and non-irrigated plots

The alpha-diversity of leaf microbial communities was estimated with the Shannon index using the diversity function of the vegan package v2.5-7 (Oksanen et al., 2019). It was analysed for each tree species with a generalized linear model (GLM) with a Gamma distribution using the glm function of the stats package v4.2.1 (R Core Team, 2020). The models had the number of reads per sample, the sampling month, the irrigation treatment and the interaction between the month and irrigation treatment as fixed effects. In the case of *Q. ilex* and *P. pinaster*, the models also included leaf age (young vs. mature) as a fixed effect. The experimental block was not introduced as a random factor because it led to singular models, suggesting overfitting. A *post hoc* Wilcoxon rank-sum test was then used to compare alpha-diversity values between irrigated and non-irrigated trees, for each sampling month and each tree species taken separately, using the wilcox.test function (R base).

Dissimilarities in community composition among leaf samples were estimated using the Bray–Curtis distance after Hellinger standardization, using the vegdist and decostand functions of the vegan package v2.5-7 (Oksanen et al., 2019). A Principal Coordinate Analysis (PCoA) analysis was performed for each tree species using the pcoa function of the ape package v5.5 (Paradis & Schliep, 2019). The factors structuring the phyllosphere microbiota composition were assessed for each tree species with a Permutational multivariate analysis of variance (PERMANOVA), using the adonis2 function from the vegan package v2.5-7 (Oksanen et al., 2019). The models had the number of reads per sample, the sampling month, the leaf age (only for *Q. ilex* and *P. pinaster*), the irrigation treatment and the interaction between the month and irrigation treatment as fixed effects.

2.5.3 | Machine-learning of irrigation treatment from phyllosphere microbiota data

We used the Random Forest (RF) algorithm to learn the irrigation treatment from phyllosphere microbiota data with the ranger package v0.13.1 (Wright & Ziegler, 2017). The algorithm was applied to each tree species separately, so that the RF input data consisted in 96 trees (12 trees × 4 irrigated blocks, and 12 trees × 4 non-irrigated blocks). To begin with, all leaf samples were included in the learning, without introducing information on leaf age (young or mature leaf) and on sampling month (May or July). We excluded samples for which we separated the epiphytic and endophytic communities.

For each tree species, we created 60 ASV tables representing different dimensions of the information contained in the microbial community data. The 60 ASV tables corresponded to 3 microbial targets (fungi only, bacteria only, fungi and bacteria together) × 2 filtering thresholds (all ASVs, only abundant or prevalent ASVs) × 5 taxonomic aggregation levels (ASV, genus, family, class, order) × 2 data types (quantitative data or presence/absence data). The taxonomic aggregation was performed by summing the number of reads of all ASVs assigned to the same genus, family, class or order and discarding ASVs unassigned to the considered taxonomic level. In quantitative ASV tables, we defined abundant ASVs as those with a number of reads higher than the 3rd quartile of the ASV read number distribution. In the presence/absence ASV tables, we defined prevalent ASVs as those present in at least half of the samples. Finally, we either kept all samples in the datasets, or split the datasets according to the sampling month to compare the RF performance between May and July samples.

For each tree species and each ASV table, the RF training step involved 500 decision trees, and cross-validation was performed using the k-fold method with five splits of the dataset. We optimized each algorithm by using $N = 20$ different values of the *mtry* parameter of the ranger function from the ranger package v0.13.1 (Wright & Ziegler, 2017). The *mtry* parameter corresponds to the number of variables (in our case, the number of ASVs) randomly selected by the algorithm for each split of the decision trees. The values of the *mtry* parameter to be tested, noted m_i with $i = \{1, \ldots, N\}$, were calculated as a function of the number n of ASVs in the dataset (adapted from Chang et al., 2017):

$$m_i = \frac{i \times n}{N \times 2} + 1$$

Then, the ability of each algorithm to classify leaf samples collected from trees experiencing higher water deficit (i.e. growing in non-irrigated plots) was evaluated using the following metrics:

$$\text{error} = \frac{FP + FN}{TP + TN + FP + FN}$$

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

with TP the true positives, TN the true negatives, FP the false positives and FN the false negatives.

To check that the algorithm was indeed learning the effect of irrigation treatment and not that of tree spatial position within the experiment, we also performed a non-random cross validation. Irrigated and non-irrigated blocks were randomly paired, and each pair of blocks were used as the training dataset, while the remaining blocks were used as the testing dataset. Doing so, trees from the same blocks could not be part of the training and the testing dataset at the same time.

2.5.4 | Identification of microbial biomarkers candidates

Finally, in cases where the error rate was low (less than 10%), we identified which ASVs were the most important for classifying samples into irrigated versus non-irrigated treatment, by using the Gini index estimated by the ranger function. A null or negative Gini index means that the ASV is not informative for classification, while high values indicate that they are more informative. A significance test was performed to select the most important ASVs for classification, using the importance_pvalue function from the ranger package v0.13.1 (Wright & Ziegler, 2017). Each selected ASV was then compared to the list of epiphyte and endophyte ASVs and was assigned to a compartment accordingly (epiphytic, endophytic or both).

3 | RESULTS

3.1 | Effect of irrigation on tree water status

According to linear mixed-effect models (Table S2), the irrigation treatment had significant effects on predawn water potential, in interaction with sampling month, for 3 tree species (*Q. ilex*, *B. pendula* and *P. pinaster*). As expected, in May, the predawn water potential did not differ between irrigated and non-irrigated plots for all four tree species (Figure 2). In July, the predawn water potential was significantly lower in non-irrigated plots compared to irrigated plots for *Q. ilex* and *B. pendula*, according to *post-hoc* tests (Figure 2), suggesting that these two species experienced a higher level of water deficit in non-irrigated plots at the time of sampling. The same trend was observed in *P. pinaster*, although the difference was not significant according to the *post hoc* test.

3.2 | Effect of irrigation on leaf microbiota diversity and composition

The raw bacterial datasets consisted of 11.3×10^6, 2.2×10^6, 8.2×10^6 and 5.2×10^6 demultiplexed 16S rRNA gene reads for *Q. ilex*, *Q. robur*, *B. pendula* and *P. pinaster* respectively. The sequence filtering and ASV inference steps retained 72%, 65%, 4% and 18% of the 16S rRNA gene reads for the four tree species respectively (Table S1). The final bacterial datasets consisted of 8.1×10^6, 1.4×10^6, 3.5×10^5 and 9.6×10^5 high-quality reads distributed into 231, 150, 95 and 152 samples and grouped into 1057, 229, 76 and 120 ASVs respectively.

The raw fungal datasets consisted in 6.7×10^6, 3.6×10^6, 3.7×10^6 and 5.6×10^6 demultiplexed ITS reads for *Q. ilex*, *Q. robur*, *B. pendula* and *P. pinaster* respectively. The sequence filtering and ASV inference steps retained 45%, 60%, 76% and 71% of the ITS reads for the four tree species respectively (Table S1). The final fungal datasets consisted in 3×10^6, 2.1×10^6, 2.8×10^6 and 4×10^6 high-quality reads distributed into 232, 148, 179 and 268 samples and grouped into 943, 514, 249 and 660 ASVs respectively.

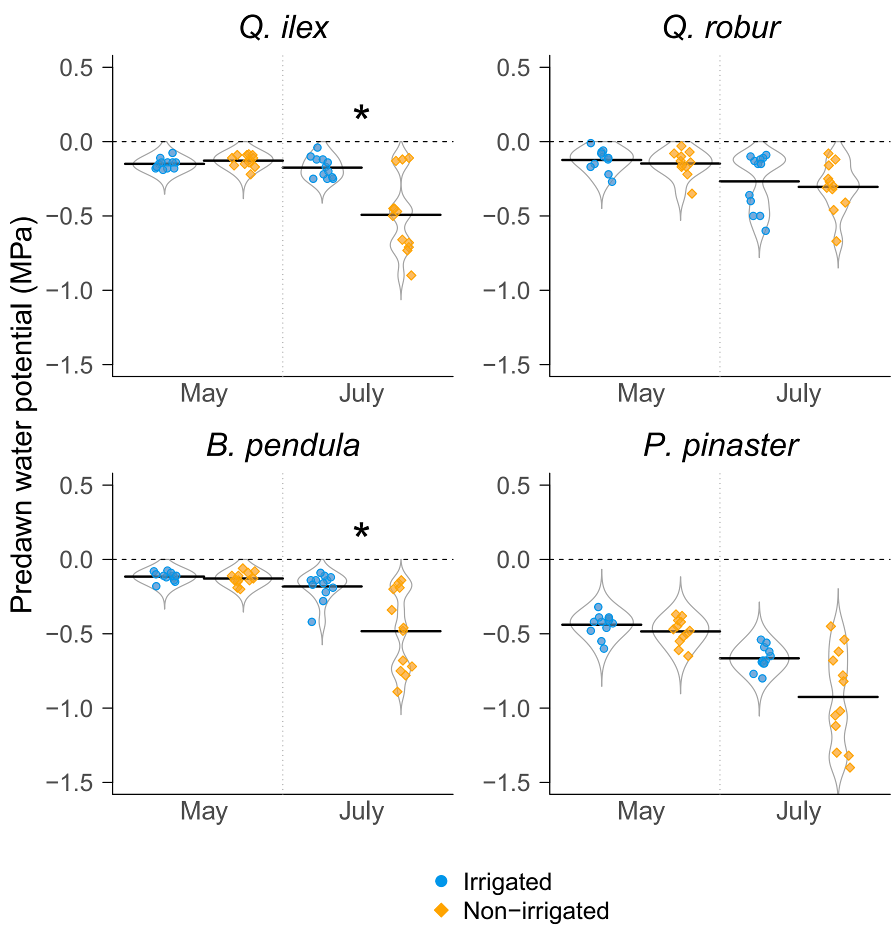

FIGURE 2 Predawn water potential of trees. Predawn water potential was measured with a pressure chamber before sunrise for 12 trees per species (3 trees per plot; Figure 1a) during each sampling campaign (May and July). Violin shapes show the distribution of water potential values and black bars represent the median. Stars indicate significant differences between irrigated and non-irrigated plots (Wilcoxon rank sum test with continuity correction, *p*-value <.05).

According to generalized linear models, the sequencing depth had a significant effect on bacterial alpha-diversity for all tree species (Table S3). The leaf age also had a significant effect for both evergreen species (*Q. ilex* and *P. pinaster*), and the sampling month had a significant effect for *Q. robur*. The *post hoc* test showed that bacterial alpha-diversity was lower in non-irrigated trees compared to irrigated one in July for *Q. robur* (Figure S1). The sequencing depth and sampling month had a significant effect on fungal alpha-diversity for all tree species but *B. pendula* (Table S3). The leaf age also had a significant effect for both evergreen species (*Q. ilex* and *P. pinaster*). Irrigation treatment also had a significant effect for *P. pinaster*, fungal-alpha diversity being slightly higher for non-irrigated trees compared to irrigated ones, although the post hoc test was not significant (Figure S1). Fungal alpha-diversity was also higher in non-irrigated trees compared to irrigated ones in July for *Q. robur* (Figure S1).

The irrigation treatment had a significant effect on bacterial beta-diversity for *Q. ilex* and *P. pinaster*, explaining 3% and 1% of variance respectively (Table S4 and Figure S2). The irrigation treatment:sampling month interaction also had a significant effect on bacterial beta-diversity for *Q. ilex*, explaining 1% of variance. The irrigation treatment had a significant effect on fungal beta-diversity for all tree species, explaining 3%, 4%, 1% and 2% of variance for *Q. ilex*, *Q. robur*, *B. pendula* and *P. pinaster* respectively (Table S4 and Figure S2). The irrigation treatment:sampling month interaction was also significant for *Q. ilex*, *Q. robur* and *B. pendula*, explaining 1%, 1% and 2% of variance respectively.

3.3 | Supervised classification of phyllosphere microbiota samples according to the irrigation treatment

After optimisation and k-fold cross validation, the RF algorithm was able to classify non-irrigated samples with 1% ± 1%, 8% ± 4%, 24% ± 6% and 21% ± 13% of error for *Q. ilex*, *Q. robur*, *B. pendula* and *P. pinaster* respectively. This optimal classification was obtained using abundant fungal ASVs for *Q. ilex*, *Q. robur*, *B. pendula* and all fungal ASVs for *P. pinaster* respectively. In general, ASV tables including fungal data gave the best results, while bacterial data alone gave higher error rates (Figure 3 and Table S5). The sensitivity, precision and error rate of the classification were not improved by aggregating the data at the genus, family, order or class level (Figure S3). The best results were obtained by using the ASV level (Figures 3 and S3 and Table S5). Almost identical results were obtained when using presence/absence ASV tables, with 2% ± 2%, 7% ± 8%, 24% ± 9% and 21% ± 10% of error rate for *Q. ilex*, *Q. robur*, *B. pendula* and *P. pinaster* respectively (Figure S4 and Table S5). The best classifications were obtained by decreasing the number of ASVs used on each branch of the classification tree

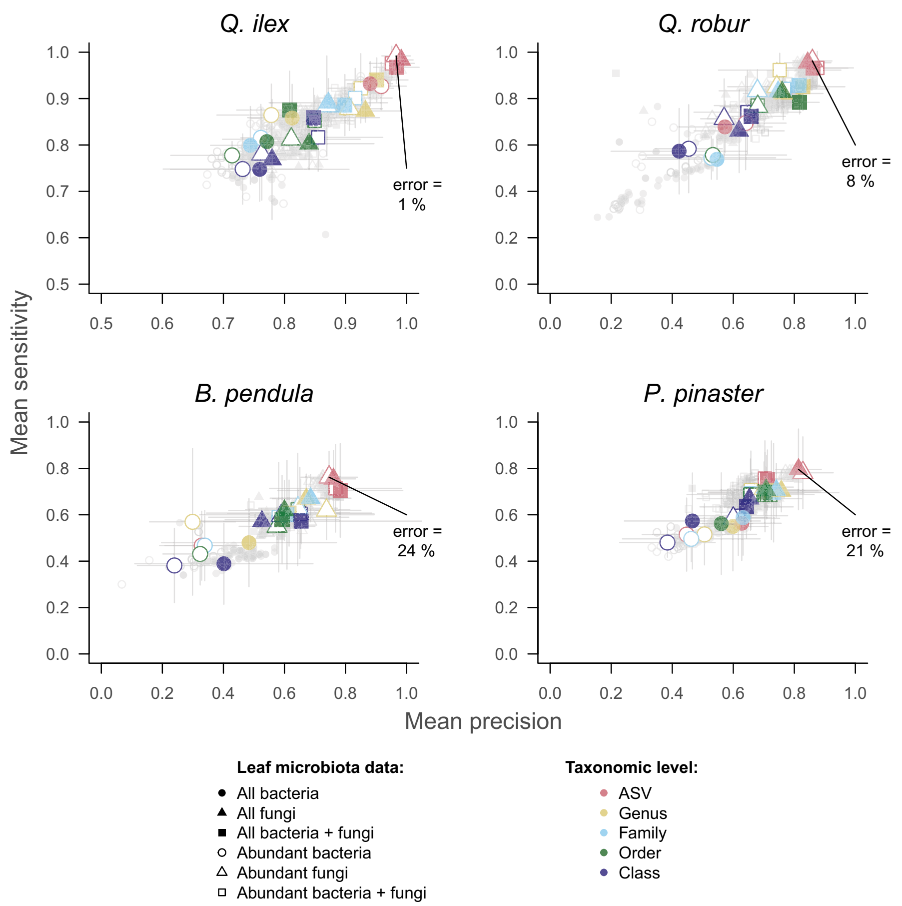

FIGURE 3 Random forest algorithm performance to classify samples from the non-irrigated treatment using quantitative phyllosphere microbiota data, depending on the microbial dataset features. Sensitivity and precision were assessed by a fivefold cross validation. Each dot represents the mean sensitivity and precision obtained for a given phyllosphere microbiota quantitative dataset and a given *mtry* parameter value. Coloured dots are obtained with the optimal *mtry* value (i.e. giving the lowest error rate). The *mtry* parameter corresponds to the number of ASVs used for each split of the decision tree. Bars represent the standard deviation over the five iterations of cross-validation. The lowest error rate is indicated in percent. Note that to improve readability, the *X*-axis for *Quercus ilex* differs from those of the other tree species.

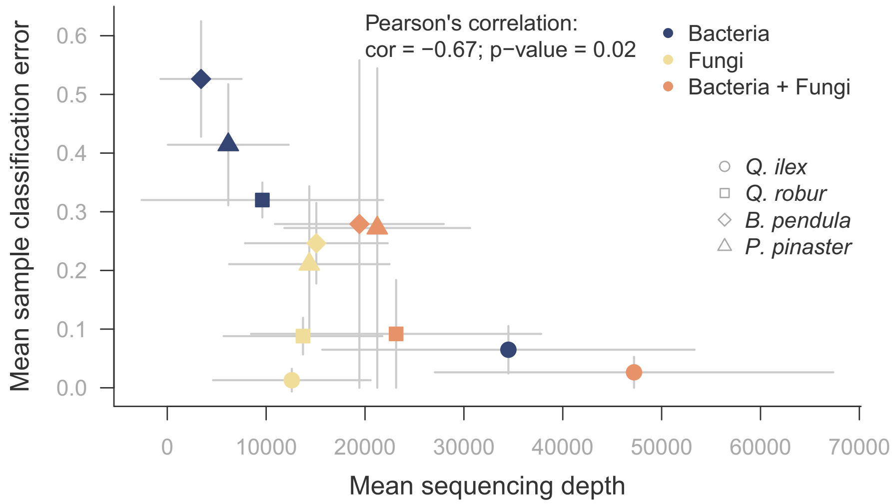

FIGURE 4 Relationship between the sequencing depth of the phyllosphere microbiota datasets and the classification error rate obtained with the Random forest algorithm using phyllosphere microbiota data. Dots represent the mean error rate obtained after fivefold cross validation for each phyllosphere microbiota dataset and each tree species, and bars represent standard deviation.

(*mtry* parameter, Table S5). Similar results were obtained using non-random cross-validation with quantitative data of abundant fungi (3% ± 3%, 9% ± 7%, 28% ± 6% and 24% ± 9% of error rate for *Q. ilex*, *Q. robur*, *B. pendula* and *P. pinaster* respectively).

We found that the error rate of classification was comparable when splitting datasets according to the sampling month (Figure S5). For *Q. ilex*, we obtained an error rate of 3% and 2% when applying the RF algorithm to fungal quantitative data from May samples only and July samples only respectively. Similarly, we obtained an error rate of 12% and 6% for *Q. robur*, 21% and 28% for *B. pendula* and 22% and 28% for *P. pinaster*.

We found that the error rate of classification was negatively correlated with the mean sequencing depth, meaning that the higher the sequencing depth, the better the classification (Figure 4, Pearson's product–moment correlation after log10 transformation of the number of reads: p-value = .02, correlation coefficient = −.67).

3.3.1 | Microbial biomarkers candidates

The most informative ASVs were analysed for *Q. ilex* and *Q. robur*, using the abundant fungi dataset, since we obtained the lowest error rates for these conditions (1% and 8% respectively, Figure 3). We identified 23 and 17 fungal ASVs significantly involved in the classification of *Q. ilex* and *Q. robur* trees respectively (Figure 5). For *Q. ilex*, the five most important fungal ASVs were assigned to the *Tremellomycetes* (ASV_59 and ASV_65), *Leotiomycetes* (ASV_6), *Paraphaeosphaeria michotii* (ASV_195) and *Dothideomycetes* (ASV_48). For *Q. robur*, the five most important fungal ASVs were assigned to the *Dothideomycetes* (ASV_48), *Cladosporium* (ASV_18), *Pseudorotiaceae* (ASV_36), *Dothideales* (ASV_28) and *Dothideomycetes* (ASV_167). Seven ASVs were important for the classification of both *Q. ilex* and *Q. robur* samples (Figure 5).

These informative ASVs were all detected in the epiphytic compartment, and sometimes in both the epiphytic and endophytic compartments. None of them was detected as a strict endophyte (Figure S6).

4 | DISCUSSION

In the present study, we investigated the effect of tree water status on the phyllosphere microbiota of four temperate tree species (*Q. robur*, *Q. ilex*, *P. pinaster* and *B. pendula*). We tested the hypothesis that the phyllosphere microbiota composition responds to tree water status and represents a biomarker that could be used in future programmes of European temperate forest ecosystems biomonitoring, in combination with other measurements of tree condition.

To test these hypotheses, we used an experimental design consisting of non-irrigated and irrigated forest plots, in which irrigation has been applied throughout the whole summer period (May to October) for several years (Castagneyrol et al., 2017). As expected, trees did not suffer from water deficit in spring according to water potential measurements, while they had a higher water deficit in non-irrigated plots compared to irrigated plots during the summer. This was true for all tree species except *Q. robur* for which the effect of irrigation treatment on summer water deficit was weaker. The experimental design thus allowed us to compare the phyllosphere microbiota of trees that have repeatedly experienced water deficit during summer (Castagneyrol et al., 2017, 2018; Galmán et al., 2022) to the microbiota of unstressed trees in irrigated plots.

4.1 | Phyllosphere microbiota responds to tree water status in European temperate forests

Leaf bacterial and fungal communities were analysed using amplicon sequencing, by considering together epiphytes and endophytes. The leaf fungal community composition (beta-diversity) varied with tree water status in the four tree species, while the bacterial community composition varied only in *P. pinaster* and *Q. ilex*. The experimental factors that we recorded in our study (sampling date, leaf age and irrigation treatment) explained a small proportion of the variance in microbial community composition, suggesting that other factors play a role. For instance, tree genotype (Faticov et al., 2021) and microclimate environment of each tree might structure the communities.

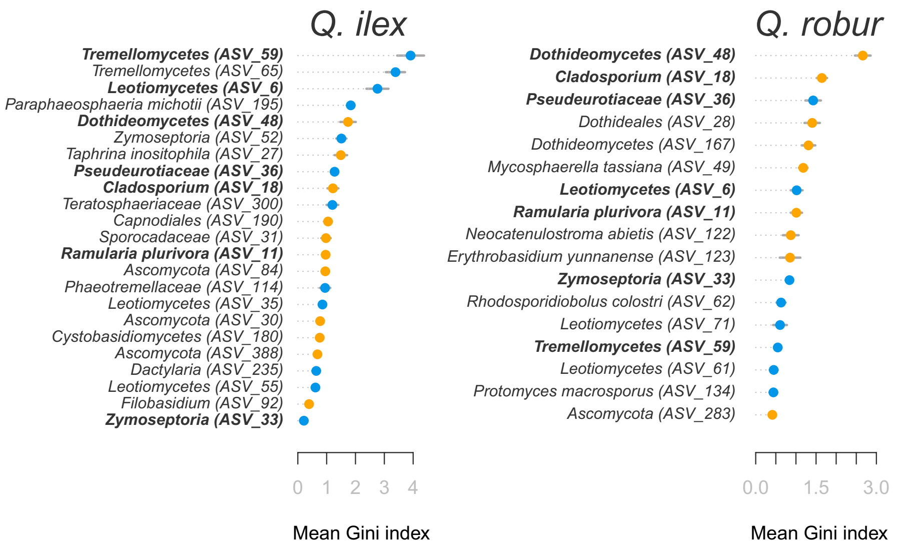

FIGURE 5 Mean Gini index and taxonomic assignment of fungal ASVs with a significant importance for classification. ASVs common to *Quercus ilex* and *Quercus robur* are in bold. Bars represent standard deviation over the 5 folds of the cross-validation in random forest classifications. Blue dots correspond to ASVs associated with the irrigated treatment, and orange dots correspond to ASVs associated with the non-irrigated treatment.

Our results are nonetheless similar to other phyllosphere studies that rarely explain a lot of variance in the microbial community composition (e.g. Debray et al., 2022; Faticov et al., 2021). The leaf fungal richness (alpha-diversity) decreased slightly with irrigation in *Q. robur* and *P. pinaster*. Interestingly, irrigation increased bacterial diversity and decreased fungal diversity in leaves of *Q. robur*, as observed recently in the tomato phyllosphere (Debray et al., 2022). Similarly, the diversity of bacterial endophytes in natural poplar populations was recently found to be reduced by drought (Firrincieli et al., 2020). The decrease in phyllosphere bacterial diversity with drought is however not a general trend, as opposite results were found in *Q. ilex* (Peñuelas et al., 2012; Rico et al., 2014). Altogether, this suggests that water deficit in trees does impact the phyllosphere microbial community, but classical community ecology analyses based on the comparison of alpha- and beta-diversity patterns might not be sufficient to extract clear and consistent patterns. In the present study, we therefore used Random Forest (RF) algorithms (Namkung, 2020) to better link phyllosphere microbiota with tree water status.

4.2 | Phyllosphere microbiota can be used as a biomarker of tree water status

Using Random Forest algorithms, we were able to classify samples according to irrigation treatments from leaf microbial community data with less than 10% error rate for *Q. robur* and *Q. ilex*, and less than 25% of error rate for *P. pinaster* and *B. pendula*. This means that we could detect trees with higher water deficit from leaf microbial communities with more than 90% of accuracy for *Q. robur* and *Q. ilex*, and more than 75% of accuracy for *P. pinaster* and *B. pendula*. We obtained similar results when making sure that trees from the same block could not be present in both the training and the testing datasets, confirming that the algorithm is learning the irrigation treatment and not the tree location within the experiment. Because *Q. robur* and *Q. ilex* trees were smaller than *P. pinaster* and *B. pendula*, the leaves were closer to the sprinkling system and leaf surface humidity may have been more influenced by the irrigation treatment, explaining a stronger response of the microbial community. Interestingly, the RF algorithms were able to predict, sometimes very accurately, the irrigation treatment of the trees, while classical PERMANOVA analyses of community composition (beta-diversity) found a very low percentage of variance explained by irrigation treatment. For instance, irrigation treatment explained 3% of the variability in *Q. ilex* fungal community, yet the fungal community contained enough information to predict irrigation treatment with 99% of accuracy using Random Forest. This result is similar to that of Wilhelm et al. (2022), who showed that Random Forest methods applied to soil bacterial community composition could accurately predict soil health properties, even though a PERMANOVA analysis showed that most variation in the bacterial community was explained by geographical location. These findings emphasize the efficiency of machine learning methods to exploit the large and complex datasets generated by the sequencing of microbial communities.

4.3 | Phyllosphere microbiota contains signature of past water deficit

Random Forest algorithms were able to predict irrigation treatments from leaf microbial community data even if leaves had not been collected during the summer water deficit period. We obtained good predictions for all four tree species using mixtures of samples collected in May (before the summer) and July (during summer), as well as mixtures of mature and new leaves for evergreen species (*Q. ilex* and *P. pinaster*). This suggests that the phyllosphere microbiota of trees is a robust biomarker of tree water status as microbial data do not need to be collected on a specific type of leaf or at a specific time. The effect of tree age on prediction accuracy should, however, be investigated as the study site was composed of trees of the same age. The prediction accuracy that we obtained with leaves sampled in May only was surprisingly high and comparable to the one obtained from leaves sampled in July only for both evergreen and deciduous species. This shows that newly emerged leaves which have never experienced water deficit can still contain a signature of the past summer water deficits and suggests that the phyllosphere microbiota could be used as a biomarker of past water deficit even in deciduous tree species. It is in agreement with another study showing that long-term seasonal patterns are the main factor influencing the phyllosphere community (Stone & Jackson, 2021). The presence of a signature of past summer water deficits in newly emerged leaves could be explained by their colonization from stem and branches microbial communities, which would keep a signature of the host physiology throughout the year. Tree physiology at the time of leaf emergence, which may depend on past stresses, could also influence directly the microbial community composition in newly emerged leaves. Hence, leaf microbial communities may not be early warning systems of environmental changes, but are good indicators of environmental conditions over the long run (Carignan & Villard, 2002).

4.4 | The best biomarkers of tree water status are fungi

For all four tree species, the best detection of tree water deficit was obtained using fungal data, compared to bacterial data. Although this could be explained by the higher sequencing depth of fungal datasets for *Q. robur*, *B. pendula* and *P. pinaster*, bacterial and fungal data were of comparable sequencing depth for *Q. ilex*, allowing for a biological comparison. In that case, fungal data predicted irrigation treatment with 99% of accuracy, as opposed to 96% of accuracy for bacterial data. This is coherent with previous studies showing contrasting effects of drought stress on phyllosphere bacterial and fungal communities (Debray et al., 2022). This also suggests that the bacterial community does not need to be sequenced to detect tree drought stress, thus reducing the cost of using phyllosphere microbiota for forest biomonitoring (Carignan & Villard, 2002).

4.5 | Microbe taxonomy is not needed to assess tree water status

We found that the ASV level was the most informative, and that the microbial data do not need to be aggregated to higher taxonomic levels to detect tree water deficit. These findings are in agreement with those of Wilhelm et al. (2022), who showed that microbial data at the ASV level are the most informative to predict soil health properties. In both studies, the bacterial and fungal ASVs can be used as biomarkers without prior knowledge on their taxonomy. These results contrast with those of Chang et al. (2017), who showed that bacterial ASVs aggregated at the order level were the most informative to predict crop productivity. This difference might be due to a higher functional redundancy in the soil microbial communities analysed by Chang et al. (2017), or to better taxonomic assignments for those communities. In our study, many ASVs could not be taxonomically assigned because the microbial sequences were not represented in reference databases. Therefore, there was a loss of information when aggregating the data based on taxonomy, because the unassigned ASV had to be removed from the data.

4.6 | Presence-absence data are sufficient to assess tree water status

Our results also showed that the signature of tree water deficit is contained in the presence of some abundant ASVs. Removing rare ASVs did not decrease the accuracy of water deficit detection. Neither did the removal of abundance information, by transforming the microbial community data into presence–absence data. These observations suggest that tree water deficit triggers a turnover within the abundant members of the phyllosphere community. As the abundance data obtained from Illumina sequencing are only semi-quantitative (e.g. Castaño et al., 2020 a recent study on ITS-like markers), they might add some noise in the data and impair water deficit detection. Our results suggest that presence-absence data of abundant ASVs are more reliable, and are sufficient to detect water deficit.

4.7 | We found fungal biomarkers of tree water status in oaks

To go further, we analysed the ASVs that were the most informative for the detection of water deficit. We focused on the two oak species (*Q. ilex* and *Q. robur*) since we obtained the best detections for these two tree species, compared to *P. pinaster* and *B. pendula*. Based on the Random Forest results, we identified seven fungal ASVs important for tree water status assessment in both oak species: ASV_6 (taxonomically assigned to *Leotiomycetes*), ASV_33 (*Zymoseptoria*), ASV_36 (*Pseudeurotiaceae*) and ASV_59 (*Tremellomycetes*) were associated with lower water deficit, while ASV_11 (*Ramularia plurivora*), ASV_18 (*Cladosporium*) and ASV_48 (*Dothideomycetes*) were associated with higher water deficit. All those fungal ASVs were detected both in the total leaf community (obtained by sequencing ground leaf discs) and in the epiphytic community (obtained by swabbing leaf surfaces), suggesting that they are all epiphytes, with an endophytic stage for some of them (obtained by sequencing ground leaf discs after surface sterilization). These results also suggest that swabs could be used to sample the leaf microbial community, instead of sampling full leaves and making discs. This is encouraging for the practicality of using the phyllosphere microbiota as a biomarker of tree water status in the future. Some recently developed tools such as LANDMark (Rudar et al., 2022) have been shown to outperform Random Forests in the identification of biomarkers, and could be used on our dataset to confirm the stress marker potential of the seven fungal ASVs we identified in this study.

4.8 | Sequencing data quality influences tree water deficit detection

The use of phyllosphere microbiota as a biomarker of tree water status requires high quality sequencing data. In our study, extracting high quality and quantity of DNA has been more challenging for *B. pendula* and *P. pinaster* than for the two oak species. This difficulty, combined with the necessity to use degenerated primers to avoid chloroplast amplification for the 16S rRNA gene, resulted in poor amplification and thus low number of bacterial reads for *B. pendula* and *P. pinaster*. This impeded the accuracy of water deficit detection for those tree species, indicating that extraction protocols may need to be optimized for each tree species to use the phyllosphere microbiota as a biomarker of tree water status in the future. As a result, the use of leaf microbial data to monitor water status in trees might still be challenging for some tree species, and its potential high-throughput and systematic implementation will rely on future technological advances. However, the speed at which molecular methods evolve and improve gives no doubt about the development of performant and cost-effective methods in the near future, which could be directly implemented in the field (Pomerantz et al., 2022).

4.9 | Training datasets are needed to detect water deficit

In addition to good quality sequencing data, detecting water deficit stress from phyllosphere microbiota using Random Forest algorithms requires training datasets describing the microbiota of trees of known water status. Building these datasets is difficult because it requires predawn water potential measurements on a large number of trees. To overcome this difficulty, future studies could assess the performance of the algorithm trained on our data to detect tree water deficit of the same oak species but in other European temperate forest sites. If detections are accurate, training datasets from studies such as ours could be used as a public resource to build classifiers, which could be subsequently used as a tool by forest practitioners in different sites. However, the performance of our method should first be tested for other climates and forest types.

4.10 | A promising tool for in situ automated biomonitoring

Overall, our results support the idea that environmental DNA and supervised machine learning methods can be efficiently combined for the next-generation biomonitoring of ecosystems (Bohan et al., 2017; Cordier, Lanzén, et al., 2018). We showed that phyllosphere microbiota sequencing data can be analysed with Random Forest algorithms to accurately assess tree water status in forest ecosystems. The fact that only presence-absence data of prevalent and epiphytic fungi were necessary to achieve a good detection of tree water deficit shows the potential for cheap and high-throughput implementation of these methods. Emerging molecular techniques for in situ detection of species, such as on-site sequencing methods nanopore (Pomerantz et al., 2022) or detection of biomarker species using ultrarapid mobile qPCR (Doi et al., 2021), could help, in the near future, to use leaf fungi as a biomarker of water deficit. The combination of in situ sequencing of fungal communities, cloud-based automated classification of sequence data, together with the use of other environmental DNA biomarkers and remotely sensed data would definitely fit in the scope of next-generation biomonitoring of ecosystems (Bohan et al., 2017). Such techniques would also fit in global and interdisciplinary monitoring programmes of forest health (Hartmann et al., 2018).

AUTHOR CONTRIBUTIONS

CV, DB and SD had the original idea of the project, secured the funding and designed the initial sampling plan. MC planned and executed the field sampling, managed the sample collection, analysed the sequencing data, wrote the R package, performed the analysis and wrote the initial draft of the manuscript. YR, IVH and CV participated in leaf sampling. RB and SD planned and executed the water potential measurements. MT and LP performed the DNA extractions, PCR amplification and sequencing. EC, EG and GLP provided advice, methods and tools for molecular biology. SR and BC gave feedback on the analysis and manuscript writing. CV and ILK provided significant inputs on the manuscript, and all authors revised the manuscript.

ACKNOWLEDGEMENTS

The authors thank INRAE—UEFP (https://doi.org/10.15454/1.5483264699193726E12) for the installation and management of the ORPHEE tree diversity experimental site, the PHENOBOIS facility for ecophysiological measurements, the PGTB sequencing facility (https://doi.org/10.15454/1.5572396583599417E12) for the library preparation, sequencing and demultiplexing, and the reviewers and editor for their insightful comments and help in improving the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

FUNDING INFORMATION

This work was supported by the French National Research Agency (ANR) under the grant ANR-17-CE32-0011 (NGB). Additional funding was received from the ANR (ANR-16-CE32-0003-01 to BC and IVH; ANR-10-LABX-25-01 to CV, MC and SD) and the Conseil Regionald'Aquitaine (2016-1R20301-00007218 to CV and SD).

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://doi.org/10.6084/m9.figshare.24138579 and https://doi.org/10.6084/m9.figshare.24138558.

DATA AVAILABILITY STATEMENT

The raw sequence data were deposited in the European Nucleotide Archive (ENA) under the study accession number PRJEB40190 (https://www.ebi.ac.uk/ena/browser/view/PRJEB40190; Cambon et al., 2020). The code used to process sequence data, the processed sequence files and metadata can be found at https://gitlab.com/marccamb/ngb_sequence_processing_formatting/ and on Figshare under the DOI 10.6084/m9.figshare.24138579. The Rmarkdown document, the code and result files allowing to reproduce all analysis and figures presented in the paper can be found at https://gitlab.com/marccamb/ngb_data_analysis.git and on Figshare under the DOI 10.6084/m9.figshare.24138558. The code used to aggregate ASV tables based on taxonomy, train and cross-validate the RF algorithm for several *mtry* values and to compute error rates, sensitivity and precision has been wrapped into the rf.opti.mtry function, available together with other functions used in this study in an R package (https://github.com/marccamb/microranger).

ORCID

Marine C. Cambon https://orcid.org/0000-0001-5234-4196
David A. Bohan https://orcid.org/0000-0001-5656-775X
Corinne Vacher https://orcid.org/0000-0003-3023-6113

REFERENCES

Allen, C. D., Breshears, D. D., & McDowell, N. G. (2015). On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. *Ecosphere*, *6*, art129. https://doi.org/10.1890/ES15-00203.1

Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D. D., Hogg, E. H. (Ted), Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J.-H., Allard, G., Running, S. W., Semerci, A., & Cobb, N. (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, *259*, 660–684. https://doi.org/10.1016/j.foreco.2009.09.001

Aydogan, E. L., Moser, G., Müller, C., Kämpfer, P., & Glaeser, S. P. (2018). Long-term warming shifts the composition of bacterial communities in the phyllosphere of *Galium album* in a permanent grassland field-experiment. *Frontiers in Microbiology*, *9*, 144. https://doi.org/10.3389/fmicb.2018.00144

Bálint, M., Bartha, L., O'Hara, R. B., Olson, M. S., Otte, J., Pfenninger, M., Robertson, A. L., Tiffin, P., & Schmitt, I. (2015). Relocation, high-latitude warming and host genetic identity shape the foliar fungal microbiome of poplars. *Molecular Ecology*, *24*, 235–248. https://doi.org/10.1111/mec.13018

Bohan, D. A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A. J., & Woodward, G. (2017). Next-generation global biomonitoring: Large-scale, automated reconstruction of ecological networks. *Trends in Ecology & Evolution*, *32*, 477–487. https://doi.org/10.1016/j.tree.2017.03.001

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Brodribb, T. J., Powers, J., Cochard, H., & Choat, B. (2020). Hanging by a thread? Forests and drought. *Science*, *368*, 261–266. https://doi.org/10.1126/science.aat7631

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581–583. https://doi.org/10.1038/nmeth.3869

Cambon, M. C., Trillat, M., Lesur-Kupin, I., Burlett, R., Chancerel, E., Guichoux, E., Piouceau, L., Castagneyrol, B., Le Provost, G., Robin, S., Ritter, Y., Van Halder, I., Delzon, S., Bohan, D. A., & Vacher, C. (2020). Phyllosphere microbiota from the ORPHEE experiment; European Nucleotide Archive; PRJEB40190. [Data set].

Carignan, V., & Villard, M.-A. (2002). Selecting indicator species to monitor ecological integrity: A review. *Environmental Monitoring and Assessment*, *78*, 45–61. https://doi.org/10.1023/A:1016136723584

Castagneyrol, B., Bonal, D., Damien, M., Jactel, H., Meredieu, C., Muiruri, E. W., & Barbaro, L. (2017). Bottom-up and top-down effects of tree species diversity on leaf insect herbivory. *Ecology and Evolution*, *7*, 3520–3531. https://doi.org/10.1002/ece3.2950

Castagneyrol, B., Jactel, H., & Moreira, X. (2018). Anti-herbivore defences and insect herbivory: Interactive effects of drought and tree neighbours. *Journal of Ecology*, *106*(5), 2043–2057. https://doi.org/10.1111/1365-2745.12956

Castaño, C., Berlin, A., Brandström Durling, M., Ihrmark, K., Lindahl, B. D., Stenlid, J., Clemmensen, K. E., & Olson, Å. (2020). Optimized metabarcoding with Pacific biosciences enables semi-quantitative analysis of fungal communities. *The New Phytologist*, *228*, 1149–1158. https://doi.org/10.1111/nph.16731

Chang, H.-X., Haudenshield, J. S., Bowen, C. R., & Hartman, G. L. (2017). Metagenome-wide association study and machine learning prediction of bulk soil microbiome and crop productivity. *Frontiers in Microbiology*, *8*, 519. https://doi.org/10.3389/fmicb.2017.00519

Chelius, M. K., & Triplett, E. W. (2001). The diversity of archaea and bacteria in association with the roots of Zea mays L. *Microbial Ecology*, *41*, 252–263. https://doi.org/10.1007/s002480000087

Compant, S., Clément, C., & Sessitsch, A. (2010). Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. *Soil Biology and Biochemistry*, *42*, 669–678. https://doi.org/10.1016/j.soilbio.2009.11.024

Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D. A., Bouchez, A., Chariton, A., Creer, S., Frühe, L., Keck, F., Keeley, N., Laroche, O., Leese, F., Pochon, X., Stoeck, T., Pawlowski, J., & Lanzén, A. (2021). Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. *Molecular Ecology*, *30*, 2937–2958. https://doi.org/10.1111/mec.15472

Cordier, T., Esling, P., Lejzerowicz, F., Visco, J., Ouadahi, A., Martins, C., Cedhagen, T., & Pawlowski, J. (2017). Predicting the ecological quality status of marine environments from eDNA metabarcoding data using supervised machine learning. *Environmental Science & Technology*, *51*, 9118–9126. https://doi.org/10.1021/acs.est.7b01518

Cordier, T., Forster, D., Dufresne, Y., Martins, C. I. M., Stoeck, T., & Pawlowski, J. (2018). Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. *Molecular Ecology Resources*, *18*, 1381–1391. https://doi.org/10.1111/1755-0998.12926

Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T., & Pawlowski, J. (2018). Embracing environmental genomics and machine learning for routine biomonitoring. *Trends in Microbiology*, *27*, 387–397. https://doi.org/10.1016/j.tim.2018.10.012

Cordier, T., Robin, C., Capdevielle, X., Fabreguettes, O., Desprez-Loustau, M.-L., & Vacher, C. (2012). The composition of phyllosphere fungal assemblages of European beech (*Fagus sylvatica*) varies significantly along an elevation gradient. *The New Phytologist*, *196*, 510–519. https://doi.org/10.1111/j.1469-8137.2012.04284.x

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, *6*, 226. https://doi.org/10.1186/s40168-018-0605-2

Debray, R., Socolar, Y., Kaulbach, G., Guzman, A., Hernandez, C. A., Curley, R., Dhond, A., Bowles, T., & Koskella, B. (2022). Water stress and disruption of mycorrhizas induce parallel shifts in phyllosphere microbiome composition. *The New Phytologist*, *234*, 2018–2031. https://doi.org/10.1111/nph.17817

Doi, H., Watanabe, T., Nishizawa, N., Saito, T., Nagata, H., Kameda, Y., Maki, N., Ikeda, K., & Fukuzawa, T. (2021). On-site environmental DNA detection of species using ultrarapid mobile PCR. *Molecular Ecology Resources*, *21*, 2364–2368. https://doi.org/10.1111/1755-0998.13448

Edgar, R. C., & Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, *31*, 3476–3482. https://doi.org/10.1093/bioinformatics/btv401

Faticov, M., Abdelfattah, A., Roslin, T., Vacher, C., Hambäck, P., Blanchet, F. G., Lindahl, B. D., & Tack, A. J. M. (2021). Climate warming dominates over plant genotype in shaping the seasonal trajectory of foliar fungal communities on oak. *The New Phytologist*, *231*, 1770–1783. https://doi.org/10.1111/nph.17434

Firrincieli, A., Khorasani, M., Frank, A. C., & Doty, S. L. (2020). Influences of climate on phyllosphere endophytic bacterial communities of wild poplar. *Frontiers in Plant Science*, *11*, 203. https://doi.org/10.3389/fpls.2020.00203

Galan, M., Razzauti, M., Bard, E., Bernard, M., Brouat, C., Charbonnel, N., Dehne-Garcia, A., Loiseau, A., Tatard, C., Tamisier, L., Vayssier-Taussat, M., Vignes, H., & Cosson, J.-F. (2016). 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife. *mSystems*, *1*, e00032-16. https://doi.org/10.1128/mSystems.00032-16

Galmán, A., Vázquez-González, C., Röder, G., & Castagneyrol, B. (2022). Interactive effects of tree species composition and water availability on growth and direct and indirect defences in *Quercus ilex*. *Oikos*, *2022*(5), e09125. https://doi.org/10.1111/oik.09125

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes—Application to the identification of mycorrhizae and rusts. *Molecular Ecology*, *2*, 113–118. https://doi.org/10.1111/j.1365-294X.1993.tb00005.x

Gerhardt, A. (2002). Bioindicator species and their use in biomonitoring. In *Environmental monitoring* (pp. 77–123). Encyclopedia of Life Support Systems (EOLSS).

Hacquard, S., & Schadt, C. W. (2015). Towards a holistic understanding of the beneficial interactions across the Populus microbiome. *The New Phytologist*, *205*, 1424–1430. https://doi.org/10.1111/nph.13133

Hacquard, S., Spaepen, S., Garrido-Oter, R., & Schulze-Lefert, P. (2017). Interplay between innate immunity and the plant microbiota. *Annual Review of Phytopathology*, *55*, 565–589. https://doi.org/10.1146/annurev-phyto-080516-035623

Hardoim, P. R., van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., Döring, M., & Sessitsch, A. (2015). The hidden world within plants: Ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiology and Molecular Biology Reviews*, *79*, 293–320. https://doi.org/10.1128/MMBR.00050-14

Hartmann, H., Schuldt, B., Sanders, T. G. M., Macinnis-Ng, C., Boehmer, H. J., Allen, C. D., Bolte, A., Crowther, T. W., Hansen, M. C., Medlyn, B. E., Ruehr, N. K., & Anderegg, W. R. L. (2018). Monitoring global tree mortality patterns and trends. Report from the VW symposium 'crossing scales and disciplines to identify global trends of tree mortality as indicators of forest health.' *The New Phytologist*, *217*, 984–987. https://doi.org/10.1111/nph.14988

Kembel, S. W., O'Connor, T. K., Arnold, H. K., Hubbell, S. P., Wright, S. J., & Green, J. L. (2014). Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 13715–13720.

Kim, H., Lee, K. K., Jeon, J., Harris, W. A., & Lee, Y.-H. (2020). Domestication of Oryza species eco-evolutionarily shapes bacterial and fungal communities in rice seed. *Microbiome*, *8*, 20.

Knights, D., Costello, E. K., & Knight, R. (2011). Supervised classification of human microbiota. *FEMS Microbiology Reviews*, *35*, 343–359. https://doi.org/10.1111/j.1574-6976.2010.00251.x

Kristy, B., Carrell, A. A., Johnston, E., Cumming, J. R., Klingeman, D. M., Gwinn, K., Syring, K. C., Skalla, C., Emrich, S., & Cregger, M. A. (2022). Chronic drought differentially alters the belowground microbiome of drought-tolerant and drought-susceptible genotypes of *Populus trichocarpa*. *Phytobiomes Journal*, *6*, 317–330. https://doi.org/10.1094/PBIOMES-12-21-0076-R

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Laforest-Lapointe, I., Messier, C., & Kembel, S. W. (2017). Tree leaf bacterial community structure and diversity differ along a gradient of urban intensity. *mSystems*, *2*, e00087-17. https://doi.org/10.1128/mSystems.00087-17

Lata, R., Chowdhury, S., Gond, S. K., & White, J. F. (2018). Induction of abiotic stress tolerance in plants by endophytic microbes. *Letters in Applied Microbiology*, *66*, 268–276. https://doi.org/10.1111/lam.12855

Liu, Y.-R., Delgado-Baquerizo, M., Bi, L., Zhu, J., & He, J.-Z. (2018). Consistent responses of soil microbial taxonomic and functional attributes to mercury pollution across China. *Microbiome*, *6*, 183.

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

McLaren, M. R., & Callahan, B. J. (2020). Pathogen resistance may be the principal evolutionary advantage provided by the microbiome. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*, 20190592. https://doi.org/10.1098/rstb.2019.0592

Namkung, J. (2020). Machine learning methods for microbiome studies. *Journal of Microbiology*, *58*, 206–216.

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., & Tedersoo, L. (2019). Mycobiome diversity: High-throughput sequencing and identification of fungi. *Nature Reviews. Microbiology*, *17*, 95–109. https://doi.org/10.1038/s41579-018-0116-y

Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Kõljalg, U., & Abarenkov, K. (2018). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research*, *47*, D259–D264. https://doi.org/10.1093/nar/gky1022

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., … Weedon, J. (2022). *vegan: Community ecology package*. R package version 2.6-4. https://CRAN.R-project.org/package=vegan

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Peñuelas, J., Rico, L., Ogaya, R., Jump, A. S., & Terradas, J. (2012). Summer season and long-term drought increase the richness of bacteria and fungi in the foliar phyllosphere of *Quercus ilex* in a mixed Mediterranean forest. *Plant Biology*, *14*, 565–575. https://doi.org/10.1111/j.1438-8677.2011.00532.x

Perreault, R., & Laforest-Lapointe, I. (2022). Plant-microbe interactions in the phyllosphere: Facing challenges of the anthropocene. *The ISME Journal*, *16*, 339–345. https://doi.org/10.1038/s41396-021-01109-3

Pineda, A., Kaplan, I., & Bezemer, T. M. (2017). Steering soil microbiomes to suppress aboveground insect pests. *Trends in Plant Science*, *22*, 770–778. https://doi.org/10.1016/j.tplants.2017.07.002

Pomerantz, A., Sahlin, K., Vasiljevic, N., Seah, A., Lim, M., Humble, E., Kennedy, S., Krehenwinkel, H., Winter, S., Ogden, R., & Prost, S. (2022). Rapid in situ identification of biological specimens via DNA amplicon sequencing using miniaturized laboratory equipment. *Nature Protocols*, *17*, 1415–1443. https://doi.org/10.1038/s41596-022-00682-x

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*, D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rho, H., Van Epps, V., Wegley, N., Doty, S. L., & Kim, S.-H. (2018). Salicaceae endophytes modulate stomatal behavior and increase water use efficiency in rice. *Frontiers in Plant Science*, *9*, 188. https://doi.org/10.3389/fpls.2018.00188

Rico, L., Ogaya, R., Terradas, J., & Peñuelas, J. (2014). Community structures of N2-fixing bacteria associated with the phyllosphere of a Holm oak forest and their response to drought. *Plant Biology*, *16*, 586–593. https://doi.org/10.1111/plb.12082

Rodriguez, R. J., Henson, J., Van Volkenburgh, E., Hoy, M., Wright, L., Beckwith, F., Kim, Y.-O., & Redman, R. S. (2008). Stress tolerance in plants via habitat-adapted symbiosis. *The ISME Journal*, *2*, 404–416. https://doi.org/10.1038/ismej.2007.106

Rosado, B. H. P., Almeida, L. C., Alves, L. F., Lambais, M. R., & Oliveira, R. S. (2018). The importance of phyllosphere on plant functional ecology: A phyllo trait manifesto. *The New Phytologist*, *219*, 1145–1149. https://doi.org/10.1111/nph.15235

Rudar, J., Porter, T. M., Wright, M., Golding, G. B., & Hajibabaei, M. (2022). LANDMark: An ensemble approach to the supervised selection of biomarkers in high-throughput sequencing data. *BMC Bioinformatics*, *23*, 110. https://doi.org/10.1186/s12859-022-04631-z

Stone, B. W. G., & Jackson, C. R. (2021). Seasonal patterns contribute more towards phyllosphere bacterial community structure than short-term perturbations. *Microbial Ecology*, *81*, 146–156. https://doi.org/10.1007/s00248-020-01564-z

Turner, S., Pryer, K. M., Miao, V. P. W., & Palmer, J. D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *The Journal of Eukaryotic Microbiology*, *46*, 327–338. https://doi.org/10.1111/j.1550-7408.1999.tb04612.x

UNITE Community. (2019). *UNITE general FASTA release for Fungi 2. Version 18.11.2018*. UNITE Community.

Unterseher, M., & Schnittler, M. (2009). Dilution-to-extinction cultivation of leaf-inhabiting endophytic fungi in beech (*Fagus sylvatica* L.)—Different cultivation techniques influence fungal biodiversity assessment. *Mycological Research*, *113*, 645–654. https://doi.org/10.1016/j.mycres.2009.02.002

Vacher, C., Castagneyrol, B., Jousselin, E., & Schimann, H. (2021). Trees and insects have microbiomes: Consequences for forest health and management. *Current Forestry Reports*, *7*, 81–96. https://doi.org/10.1007/s40725-021-00136-9

Vacher, C., Hampe, A., Porté, A. J., Sauer, U., Compant, S., & Morris, C. E. (2016). The phyllosphere: Microbial jungle at the plant–climate interface. *Annual Review of Ecology, Evolution, and Systematics*, *47*, 1–24. https://doi.org/10.1146/annurev-ecolsys-121415-032238

Vannier, N., Agler, M., & Hacquard, S. (2019). Microbiota-mediated disease resistance in plants. *PLoS Pathogens*, *15*, 1–7. https://doi.org/10.1371/journal.ppat.1007740

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*, 5261–5267. https://doi.org/10.1128/aem.00062-07

White, T., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. Innis, D. Gelfand, J. Sninsky, & T. White (Eds.), *PCR protocols: A guide to methods and applications* (pp. 315–322). Academic Press. https://doi.org/10.1016/b978-0-12-372180-8.50042-1

Wilhelm, R. C., van Es, H. M., & Buckley, D. H. (2022). Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biology and Biochemistry*, *164*, 108472. https://doi.org/10.1016/j.soilbio.2021.108472

Wright, M., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*, 1–17. https://doi.org/10.18637/jss.v077.i01

Xu, N., Zhang, Z., Shen, Y., Zhang, Q., Liu, Z., Yu, Y., Wang, Y., Lei, C., Ke, M., Qiu, D., Lu, T., Chen, Y., Xiong, J., & Qian, H. (2022). Compare the performance of multiple binary classification models in microbial high-throughput sequencing datasets. *Science of the Total Environment*, *837*, 155807. https://doi.org/10.1016/j.scitotenv.2022.155807

Zhu, Y.-G., Xiong, C., Wei, Z., Chen, Q.-L., Ma, B., Zhou, S.-Y.-D., Tan, J., Zhang, L.-M., Cui, H.-L., & Duan, G.-L. (2022). Impacts of global change on the phyllosphere microbiome. *The New Phytologist*, *234*, 1977–1986. https://doi.org/10.1111/nph.17928

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Cambon, M. C., Trillat, M., Lesur-Kupin, I., Burlett, R., Chancerel, E., Guichoux, E., Piouceau, L., Castagneyrol, B., Le Provost, G., Robin, S., Ritter, Y., Van Halder, I., Delzon, S., Bohan, D. A., & Vacher, C. (2023). Microbial biomarkers of tree water status for next-generation biomonitoring of forest ecosystems. *Molecular Ecology*, *00*, 1–15. https://doi.org/10.1111/mec.17149