INRAΣ

Annals of Forest Science

DATA PAPER

Leaf microbiome data for European beech (*Fagus sylvatica*) at the leaf and canopy scales collected in a gallery forest in South-West France

Tania Fort[1,2], Charlie Pauvert[1,3], Emilie Chancerel[1], Regis Burlett[1], Lisa Wingate[4] and Corinne Vacher[1*]

Key message

The datasets describe bacterial and fungal communities of European beech (*Fagus sylvatica*) leaves collected along a vertical gradient in a gallery forest throughout the growing season. They also describe communities in the surrounding environment of beech trees. Dataset access is at https://doi.org/10.7910/DVN/FFHAQU, and associated metadata are available at https://metadata-afs.nancy.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/f17fe848-fc3e-4297-be11-9871b35a1be4. Both can be used to uncover the dynamics and assembly processes of phyllosphere microbial communities in forest ecosystems.

Keywords Microbial ecology, Forest microbiome, Fungi, Bacteria, Phyllosphere, *Fagus sylvatica*

1 Background

Leaves are colonized by an enormous diversity of microbial species that influence tree health and growth as well as nutrient cycles in forest ecosystems (Laforest-Lapointe et al. 2017; Terhonen et al. 2019). The diversity and composition of these microbial communities are shaped by the selection exerted by leaf traits and microclimate (Vacher et al., 2016). Both factors vary across the leaf and within the tree canopy and shape the spatial structure of phyllosphere microbial communities (PMCs), thus

modulating their role in ecosystem functioning (Cordier et al. 2012; Stone & Jackson 2019; Bahram et al. 2022).

The objectives of the study were to quantify the magnitude of within-leaf and within-canopy variations in PMCs and to investigate the consequences of these spatial variations on microbial succession during leaf decomposition, using a metabarcoding approach. Data were collected on three adult beech trees at three dates during the growing season. Spatial variations in PMCs were quantified by collecting leaves at several positions within the canopy and several positions within the leaf (main vein, blade edge and blade central part, leaf surface and internal tissue). Community trajectories during leaf decomposition were investigated by incubating leaves collected on the two most distant positions of the canopy during 2 months (Fig. 1).

2 Methods

2.1 Study site and sampling design

Samples were collected in the canopy and the surrounding environment of three European beech (*Fagus sylvatica*) trees growing in the Ciron Valley forest (44° 23′ 2″ North,

---

BMC

0° 18′ 21″ West), situated approximately 50 km South-East of Bordeaux, France. The trees were 14 to 55 m from the Ciron River and 200 to 400 m from each other (Fig. 1A). Samples were collected on three sampling dates (D1, D2, and D3) corresponding to the beginning, middle, and end of the growing season (30th to 31st of May, 4th to 5th of July, and 19th to 20th of September 2017, respectively). During sampling campaigns, six southwest facing branches were collected from each tree by tree climbers, without directly touching the leaves with hands. Branches were collected at three heights, corresponding to the top, middle, and bottom of the crown. At each height, one branch was sampled from the inner part of the crown and another from the outer part (Fig. 1B). Six positions within each crown were therefore sampled (IB for Bottom of the Inner part of the crown; IM, Middle of the Inner crown; IH, Highest branch of the Inner crown; EB, Bottom of the External part of the crown; EM, Middle of the External crown; EH, Highest branch of the External crown). Branches were immediately brought back to the field lab that was set up near the three trees to process the leaves for phyllosphere microbial community analyses.

2.2 Leaf processing for the analysis of within-canopy variations in microbial communities—Set#1

On each branch, four leaves were randomly chosen (Fig. 1C), collected using sterilized nitrile gloves and placed on autoclaved filter papers. Each leaf was then folded with sterile pliers, placed in a 20-mL high-density polyethylene (HDPE) vial (Zinsser Analytic®) and frozen with liquid nitrogen. Vials were then stored in a nitrogen dry shipper (Voyageur, Air liquide®). Back in the lab, samples were cold-ground using a GenoGrinder (30 s at 1750 RPM), after adding two autoclaved stainless steel beads to the tubes. Aliquots of 20 mg of powder were then transferred to individual tubes (Micronic, MP32022, In Vitro A/S, Fredensborg, Denmark) along with two sterile stainless steel beads (TissueLyser II, Retsch, Qiagen®) and stored at − 80 °C until DNA extraction. These 216 leaf samples are hereafter referred to as Set#1.

2.3 Leaf processing for the analysis of within-leaf variations in microbial communities—Set#2 and Set#3

Three leaves were collected from the two most distant branches (cut from the lower inner part and higher outer part of the crown, respectively) to assess community variations between leaf blade and midrib. Leaves were collected with pliers and a pair of scissors cleaned with DNaway and 70% alcohol between each use. On each leaf, two 1 cm × 3 cm pieces of leaf material were cut, one at the edge of the leaf blade and another at the center of the leaf blade, taking care to avoid the midrib. The entire midrib was also removed and cut into 3 to 4
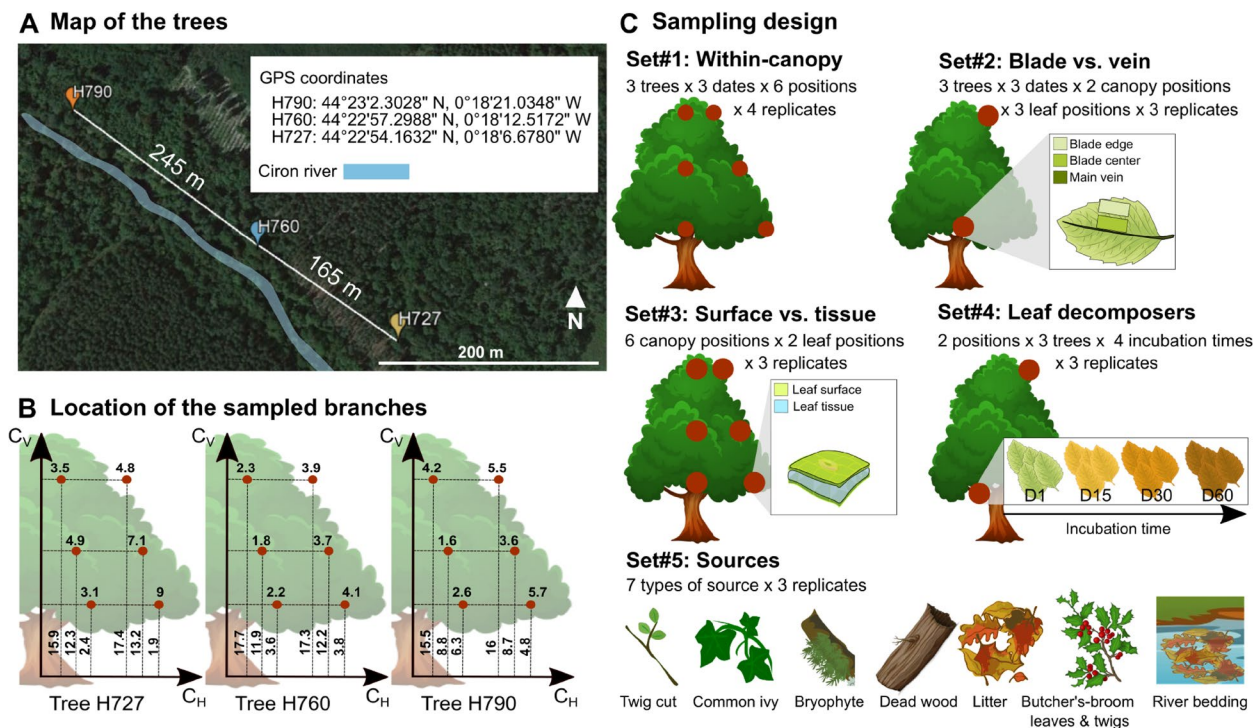
Fig. 1 Sampling design. **A** Map of the three European beech (*Fagus sylvatica*) trees sampled. **B** Position of the sampled branches in each tree. Horizontal values represent the distance to the trunk (C_H, in meters) and vertical values represent the height of the sampled branch (C_V, in meters). **C** Sampling design for each metabarcoding dataset (Set#1 to Set#5).

pieces (Fig. 1C). The samples were then stored in Eppendorf tubes. Samples were transported in a dry-shipper and stored at − 80 °C until further analysis. These 162 leaf samples are hereafter referred to as Set#2.

Three leaves per branch were also collected from 6 branches to assess community variations between the leaf surface (epiphytic communities) and the internal tissues (endophytic communities) (Fig. 1C). Each leaf was collected with sterile nitrile gloves and placed in a sterile 50 ml Falcon tube placed in a cooler. Leaves were stored in the lab at − 20 °C until the epiphytic fractions could be detached by sonication and vortexing (sonication 3 min, vortex 5 s, shake 10 s), after adding 20 mL of PBS-Tween 20 in each tube under a laminar flow hood. The resulting solution was then centrifuged for 15 min at 4000 *g*. The supernatant was removed using a PipetBoy (Integra Biosciences, Fernwald, Germany), leaving a pellet and approximately 2 ml of solution. Solutions from one of the replicates from each tree and date collected from the same position in the canopy were resuspended, pooled, and centrifuged a second time. The second and third replicates were processed in the same manner, resulting in three replicates, each representing the same position in the canopy but different trees and dates. The supernatant was removed, resulting in 800 μL of solution. The pellet was transferred into a 1.1-mL Micronic tube (Nova-Zine, Lyon, France, ref: MP32033L) itself placed in a 2-ml Eppendorf tube. The latter was centrifuged for 5 min at 10,000 *g* and the supernatant removed. Two stainless steel beads were added to each sample and stored at − 80 °C until DNA extraction. To recover the endophyte fraction of beech leaves, leaves were removed from each Falcon tube after sonication and surface-sterilized by immersion for 3 min in a 70% ethanol solution, immersion for 2 min in a 3% calcium hypochlorite solution, and rinsing with DNAway and sterilized water. After drying on sterilized filter papers, leaves were cold-ground in Zinsser vials (as previously described). These 36 leaf samples are hereafter referred to as Set#3.

2.4 Leaf processing for the analysis of decomposer communities—Set#4
On the last sampling date, leaves were collected on each tree to characterize the succession of microbial communities during leaf decomposition. Twelve leaves were collected from two branches, cut from the lower inner part and higher outer part of the canopy, respectively (Fig. 1C). Each leaf was placed in a sterile Petri Dish with 2 mL of sterile water and incubated in a climatic chamber at 17 °C with 85.9% relative humidity, corresponding to the mean climatic conditions recorded at the sampling site in September at the field site. Four leaves were then cold-ground in Zinsser vials (as previously described)

1 day, 15 days, 30 days, and 60 days after sampling for each branch and tree sampled. Leaf powders from the same branch on the same tree were pooled resulting in three replicates per branch and tree. These 72 leaf samples are hereafter referred to as Set#4.

2.5 Environmental sampling for the detection of putative sources of phyllosphere microbiota—Set#5
Seven types of environmental samples were collected to characterize the microbial communities in the environment of each tree: (1) twigs cut from the branches collected in the canopy, (2) common ivy (*Hedera helix* L.) and (3) bryophytes growing on the tree trunk, (4) dead wood bark, (5) litter and (6) Butcher's broom leaves and twigs (*Ruscus aculeatus*) beneath each tree, and (7) river bedding. All samples were collected in triplicates with sterile latex gloves and stored in 50 mL Falcon tubes, transported in coolers, and stored at − 20 °C until the epiphytic microorganisms could be detached by sonication and vortexing (as previously described, except that the first solution was filtered with a sterile cell strainer (100 μm of porosity, nylon; Dutscher) to remove bark and litter fragments). These 21 plant samples are hereafter referred to as Set#5.

2.6 Metabarcoding of bacterial and fungal communities
DNA extraction and amplification were performed under a hood in a confined laboratory. Total genomic DNA of leaf and environmental samples was extracted using the Qiagen DNeasy plant kit (Qiagen) accordingly to the manufacturer's protocol, except that samples were incubated with AP1 solution for 1 h at 65 °C and DNA extracts were eluted twice, the first time with 50 μL of Buffer APE and the second time with the first elution solution.

The V5-V6 region of the bacterial 16S rRNA gene was amplified using 16S primers 799F-1115R (Redford et al., 2010; Chelius and Triplett, 2001) to exclude chloroplast DNA. To avoid a two-stage PCR protocol and reduce sequencing biases, each primer contained the Illumina adaptor sequence, a tag, and a heterogeneity spacer, as described in Laforest-Lapointe (2017) (799F: 5′-CAA GCAGAAGACGGCATACGAGATGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCTxxxxxxxxxxxxHS-AACMGGATTAGATACCCKG-3′; 1115R: 5′-AATGAT ACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCTxxxxxxxxxxxxHS-AGG GTTGCGCTCGTTG-3′). The PCR mixture (20 μL of final volume) consisted of 2 μL of each of the forward and reverse primers (2 μM), 2 μL of dNTPs (2 mM), 4 μL of 5X HotStart Phusion HF Mix, 0.6 μL of DMSO, 0.2 μL of Phusion Hot Start II polymerase (ThermoScientific), 1 μL of template, and water up to 20 μl. PCR cycling

reactions were conducted on a Veriti 96-well Thermal Cycler (Applied Biosystems) using the following conditions: initial denaturation at 98 °C for 30 s followed by 35 cycles at 98 °C for 15 s, 60 °C for 30 s, 72 °C for 30 s with final extension of 72 °C for 10 min. For 16S primers, HS represented a 0–7-base-pair heterogeneity spacer and for all primers "x" a 12 nucleotides tag.

The fungal ITS1 region of the nuclear ribosomal internal transcribed spacer, considered the universal barcode marker for fungi (Schoch et al. 2012), was amplified using the ITS1F (5′- CAAGCAGAAGACGGCATACGA GATGTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTxxxxxxxxxxxx-3′) and ITS2 (5′-AATGATACG GCGACCACCGAGATCTACACTCTTTCCCTACAC GACGCTCTTCCGATCTxxxxxxxxxxxx-3′)   primers. The PCR mixture (20 μL of final volume) consisted of 1 μL of each of the forward and reverse primers (2 μM), 4 μL of 5X HotFirePol Mix (Solis BioDyne), 3 μL of template, and water up to 20 μl. PCR cycling reactions were conducted on a Veriti 96-well Thermal Cycler (Applied Biosystems) using the following conditions: initial denaturation at 95 °C for 15 min followed by 35 cycles at 95 °C for 15 s, 55.8 °C for 30 s, 72 °C for 45 s with final extension of 72 °C for 10 min.

Each PCR plate had 2 positive controls. Bacterial positive controls were represented by the DNA of two marine bacterial strains (*Sulfitobacter pontiacus* and *Vibrio splendidus*); the first positive control included 1 μL of 10 ng μL^{-1} DNA of *Vibrio splendidus* only, and the second included 1 μL of an equimolar mixture of both strains. Fungal positive controls were represented by the DNA of one fungal strain growing on hyper saline environments (*Debaryomyces hansenii*) and a second strain growing on hyper sweet or salty environments (*Wallemia sebi*); the first positive control included 3 μL of 10 ng μL^{-1} DNA of *Debaryomyces hansenii* only, and the second included 3μL of an equimolar mixture of both strains.

Each PCR plate also had several negative controls. Negative PCR controls were represented by at least one well containing PCR mix without any DNA template. Negative extraction controls were represented by at least four wells containing PCR mix with DNA extract resulting from the extraction of empty collection microtubes. Four additional negative controls were also included: (1) potential contamination during leaf weighing was assessed by opening three empty tubes during this time window and rinsing them with 1 mL of sterile water; (2) potential contamination during leaf grinding was analyzed in the same way; (3) the putative contaminants present in the PBS-Tween solution used to separate the epiphytic and endophytic communities were analyzed

in approximately 300 μL of the solution; and (4) putative contaminants present in the sterile water added to the Petri dishes during the decomposition of the leaves were also analyzed in 300 μL of the solution. Each type of negative control was represented by 1 ml of solution of 3 replicates pooled together and then extracted.

Amplifications were confirmed by electrophoresis on a 2% agarose gel. PCR products were purified, quantified (Quant-it dsDNA assay kit; Invitrogen), and equimolarly pooled (Hamilton Microlab STAR robot). Average fragment size was checked using a Tapestation instrument (Agilent Technologies). Libraries were sequenced on four runs of the MiSeq Instrument (Illumina) with the reagent kit v2 (500-cycles). Sequence demultiplexing (with exact index search) was performed at the PGTB sequencing facility (Genome Transcriptome Facility of Bordeaux, Pierroton, France) using DoubleTagDemultiplexer.

2.7 Bioinformatic analysis

Primers of bacterial 16S and fungal ITS1 reads were removed using cutadapt (Martin 2011), and sequences were then analyzed using DADA2 v1.12.1 (Callahan et al. 2016). Only sequences with less than two expected errors and longer than 50 bp were retained. Quality reads were then assembled into Amplicon Sequence Variants (ASVs) using the dada function (with default options) and chimeric sequences were removed using the consensus method of the removeBimeras function. ASVs that differed in length but were otherwise identical (i.e., that had no mismatches or internal indels when aligned) were collapsed together using the CollapseNoMismatch function. Taxonomic assignments were performed using the RDP classifier (Wang et al. 2007) implemented in DADA2 and trained with the UNITE fungal database v8.2 (Abarenkov et al. 2018; Abarenkov et al. 2020) and SILVA bacterial database v138 (Quast et al. 2013), with an 80% confidence threshold. The ASV tables were then imported in R using the phyloseq package v1.26.0 (McMurdie & Holmes 2013) and filtered. Only ASVs assigned to a bacterial and fungal kingdom were kept. ASVs assigned to chloroplast were also removed from the bacterial amplicons dataset. Positive controls were used to remove contaminants (as described in Galan et al. 2016). The false-assignment threshold (T_{FA}) was defined as the highest sequence count of a positive control strain in a non-control sample, divided by the total number of sequences of the strain in the whole run and multiplied by the total number of sequences of each ASV. ASVs were removed from all samples where they harbored fewer sequences than the T_{FA} threshold. Additionally, the negative controls were used to identify and

remove contaminants using the *DivComAnalyses* package v0.9 (https://github.com/fconstancias/DivComAnalyses) and the *microDecon* package v1.0.2 (McKnight et al. 2019). Singletons and samples with less than 100 reads were removed from all datasets.

2.8 Statistical analysis

All statistical analyses were performed with R (R Development Core Team, 2019). Variations in microbial community composition were analyzed by using permutational multivariate analyses of variance (PERMANOVAs) with 9999 permutations, performed with the *adonis* function of the vegan package v.2.5.6 (Oksanen et al., 2019). Compositional dissimilarities among samples were estimated using a binary version of the Jaccard index (Jaccard 1901) and visualized with principal coordinate analyses (PCoAs) using the *phyloseq* package v.1.34.0 (McMurdie & Holmes 2013).

Within-canopy variations in microbial community composition were assessed using Set#1 (Fig. 1 C). The vertical (C_V) and horizontal (C_H) leaf position within the canopy (Fig. 1B), the sampling date (D), the tree (T), and all their interactions were tested as fixed effects.

Within-leaf variations in microbial community composition were assessed using Set#2 and Set#3 (Fig. 1C). Within-canopy position (C), within-leaf position (L_{BV}: blade edge, blade central part, main vein), and their interaction were tested using Set#2 with permutations constrained by date (D) and tree (T). Within-canopy position (C), within-leaf position (L_{ST}: surface vs internal tissue), and their interaction were tested using Set#3.

The effect of leaf position within the canopy on microbial community trajectory during early decomposition was assessed using Set#4 (Fig. 1C). Within-canopy position (C), incubation time (IT), and their interaction were tested with permutations constrained by tree (T).

2.9 Access to the data and metadata description

The sequences were generated with four MiSeq runs (Run 1 to 4; Table 1). Run 1 and Run 3 include, respectively, the bacterial and fungal reads characterizing within-canopy variations (Set#1; Fig. 1 and Table 1). Run 2 and Run 4 include, respectively, bacterial and fungal reads characterizing within-leaf variations (Set#2 and Set#3; Fig. 1 and Table 1), decomposing leaves (Set#4; Fig. 1 and Table 1), and putative sources of microorganisms (Set #5; Fig. 1 and Table 1).

The present data paper provides, for each MiSeq run, a FASTA file, an ASV table, and a metadata table. FASTA files include ASV identifiers and ASV sequences. ASV tables include ASV identifiers and microbial read counts obtained in each sample after filtering. The metadata tables include sampling dates, trees identifier and height, sample type, tissue type (when needed), tree distance to the river, branch position within the canopy including branch height, distance to the trunk, angle to the trunk, and total read counts before and after filtering.

Raw sequences are available from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the accession numbers PRJNA699461 (16S reads) and PRJNA699462 (ITS1

Table 1 Description of the Illumina MiSeq seqencing runs and the sequence data sets (Set#1 to Set#5 represented in Fig. 1C). For each data set, the table includes minimum (min), maximum (max), mean, median and total number of microbial reads per sample, and number of samples and Amplicon Sequence Variants (ASVs) after all quality filters. The total number of samples before quality filters is given into brackets

Run	Set	Target gene	Min reads	Mean reads	Median reads	Max reads	#Total reads	#Samples	#ASVs
Run1	Set#1	V5–V6 region of the bacterial 16S	223	27,008.08	25,427	77,777	5,509,649	204 [216]	7994
Run2	All sets	V5–V6 region of the bacterial 16S	1877	14,934.57	13,900	49,186	4,062,204	272 [291]	8582
	Set#2		2597	16,452.09	15,320	49,186	2,484,265	151 [162]	4953
	Set#3		1877	11,675.1	10,428	31,533	361,928	31 [36]	1393
	Set#4		2700	13,720.46	13,385.5	24,499	960,432	70 [72]	2649
	Set#5		3696	12,778.95	10,538	25,180	255,579	20 [21]	2725
Run 1 + Run 2	All sets		223	20,108.93	18,370.5	77,777	9,571,853	476 [507]	13,173
Run3	Set#1	Fungal ITS1 region	2555	29,429.85	24,368	90,458	6,062,549	206 [216]	5951
Run4	All sets	Fungal ITS1 region	1002	29,698.18	27,848	93,873	8,523,378	287 [291]	8700
	Set#2		3176	32,634.37	28,617	93,873	5,254,134	161 [162]	5498
	Set#3		5392	27,488.14	23,614	82,485	989,573	36 [36]	1574
	Set#4		5220	27,000.31	27,334	48,015	1,944,022	72 [72]	2736
	Set#5		1002	18,647.17	15,351.5	41,041	335,649	18 [21]	2906
Run 3 + Run 4	All sets		1002	29,586.06	26,327	93,873	14,585,927	493 [507]	11,242

reads). The R scripts for the bioinformatic analysis (Pipeline_16S, Pipeline_ITS) that generated the FASTA files, filtered ASV tables, and associated metadata are available at https://doi.org/10.7910/DVN/FFHAQU. An additional description of the data and metadata is available at https://metadata-afs.nancy.inra.fr/geonetwork/srv/fre/catalog.search#/metadata/f17fe848-fc3e-4297-be11-9871b35a1be4. The R scripts used to perform statistical analysis (AFS_R_script_datastructure) are also provided.

2.10 Technical validation
Quality filtering removed 32.17% to 54.37% of the total number of reads per run (Table 2) and generated a total of 9,571,853 16S bacterial read pairs and 14,585,927 ITS1 fungal read pairs corresponding to 13,173 bacterial and 11,242 fungal ASVs (Table 1). The percentages of reads lost for each pipeline step are described in Appendix-Method S1 and in Table 2. The number of sequences per sample after quality filtering ranged from 223 to 93,873. Bacterial and fungal communities were represented by 476 and 493 samples, respectively (Table 1).

Statistical analysis revealed a vertical and horizontal stratification of both leaf fungal and bacterial communities within the crowns of mature beech trees (Appendix Table 3 and Fig. 2). Based on the three trees sampled, within-canopy variations were greater than between-tree variations for bacterial communities, while it was the opposite for fungal communities

(Appendix Table 3 and Fig. 2). Within-canopy variations were also larger than within-leaf variation except in the case of fungal communities, whose composition varied more between the leaf surface and the internal tissues than between positions across the canopy (Appendix Table 4). Finally, our analyses showed that leaf position within the canopy influenced microbial community trajectory during leaf decomposition (Appendix Table 5 and Fig. 2).

2.11 Reuse potential and limits
The metabarcoding datasets describe the dynamics of phyllosphere bacterial and fungal communities in three mature beech trees during the growing season until leaf decomposition, as well as their spatial variations at two scales, the within-leaf scale and the within-canopy scale. Although the assessment of within-canopy variation is limited by the low number of trees (only 3), future statistical analysis could contribute to identifying the ecological processes that shape phyllosphere microbial communities in forest ecosystems and at what spatial scale they occur. The role of dispersal could, in particular, be assessed by developing metacommunity models that integrate data from the putative sources of microorganisms collected around the trees and the geographic distances among samples (Miller et al. 2018). Furthermore, the extent to which the sampled surrounding environments contribute to the community composition could be assessed using source tracking models tailored for microbial ASV (Briscoe et al. 2022). Such models would provide a better understanding of pathogen dispersal in the tree canopy and help predict the outbreak of plant diseases in forests.

Table 2 Percentage of reads lost at each stage of the bioinformatic pipeline and over the entire bioinformatic process, for each Illumina MiSeq run

Bioinformatic pipeline step	Bacteria		Fungi	
	Run1	Run2	Run3	Run4
Primers removal with Cutadapt	11.97%	14.04%	5.12%	5.07%
Quality filtering with DADA2	7.53%	4.43%	9.47%	10.52%
Chimera removal	19.01%	19.01%	9.3%	5.76%
Removal of ASVs not assigned to a microbial kingdom	0.11%	0.12%	2.19%	0%
Decontamination (Galan's filter)	3.27%	5.23%	3.97%	3.96%
Decontamination (Decontam filter)	0.79%	11.14%	5.66%	6.86%
Removal of singletons and samples with less than 100 reads	0%	0.4%	0%	0%
Total read lost	42.68%	54.37%	35.71%	32.17%

Appendix
Method S1—Raw data curation
Bioinformatic pipelinesretained 57.3%. 45.63%. 64.29% and 67.8% of the raw MiSeq reads in Run1, Run2,Run3, and Run4, respectively (Table 2). Read losses occurred on the followingsteps: 5 to 14% of the reads lacked valid primers detected by cutadapt; 4.4 to10.5% of the reads were removed because they had more than two expected errorsaccording to DADA2 or were shorter than 50bp; chimera detection removed 5.7 to19% of the reads; 0 to 2.2% of the reads could not be assigned to a microbialkingdom and were removed; filtering the ASV table using the positive controlsremoved 3.2 to 5.2% of the sequences and negative controls removed 0.8 to 11.1%of the sequences. Filtering out singletons and samples containing less than 100reads removed 0 to 0.4% of the reads (Table 2).

Table 3 Effect of within-canopy leaf position on the composition of phyllosphere microbial communities of European beech (*Fagus sylvatica*), tested using permutational multivariate analyses of variance (PERMANOVA). Compositional dissimilarities among leaves were estimated using the binary version of Jaccard distance. The effects of sampling date (*D*), tree identity (*T*), within-canopy vertical position (C_V), and horizontal position (C_H) and their interaction were tested using sample Set#1(Fig. 1C)

	Bacteria				Fungi			
	d.f	*F*-value	R^2	*P*-value	d.f	*F*-value	R^2	*P*-value
Vertical position (C_V)	2	6.047	0.044	0.001***	2	7.192	0.048	0.001***
Horizontal position (C_H)	1	3.561	0.013	0.001***	1	3.347	0.011	0.001***
Date (D)	2	8.333	0.061	0.001***	2	7.390	0.050	0.001***
Tree (T)	2	5.441	0.040	0.001***	2	17.101	0.115	0.001***
C_V x C_H	2	1.728	0.013	0.001***	2	1.957	0.013	0.001***
C_V x D	4	2.020	0.030	0.001***	4	1.638	0.022	0.001***
C_H x D	2	2.090	0.015	0.001***	2	1.799	0.012	0.001***
C_V x T	4	2.024	0.030	0.001***	4	2.900	0.039	0.001***
C_H x T	2	1.886	0.014	0.001***	2	1.917	0.013	0.001***
D x T	4	2.233	0.033	0.001***	4	2.167	0.029	0.001***
C_V x C_H x D	4	1.538	0.023	0.001***	4	1.460	0.020	0.001***
C_V x C_H x T	4	1.659	0.024	0.001***	4	1.679	0.023	0.001***
C_V x D x T	8	1.487	0.044	0.001***	8	1.421	0.038	0.001***
C_H x D x T	4	1.446	0.021	0.001***	4	1.393	0.019	0.001***
C_V x C_H x D x T	8	1.481	0.044	0.001***	8	1.446	0.039	0.001***

Table 4 Effects of within-canopy and within-leaf positions on the composition of phyllosphere microbial communities of European beech (*Fagus sylvatica*), tested using permutational multivariate analyses of variance (PERMANOVA). Compositional dissimilarities among samples were estimated using a binary version of the Jaccard index. The effects of within-canopy position (*C*), within-leaf position (L_{BV}: blade edge, blade central part or main nerve), and their interaction were tested using sample Set#2 (Fig. 1C). The effects of within-canopy position (*C*), within-leaf position (L_{ST}: surface vs. tissue), and their interaction were tested using sample Set#3 (Fig. 1C). Permutations were constrained by sampling date (*D*) and tree identity (*T*)

	Bacteria				Fungi			
	d.f	*F*-value	R^2	*P*-value	d.f	*F*-value	R^2	*P*-value
Blade vs. main nerve								
Within-leaf position (L_{BV})	2	1.7611	0.0217	0.001***	2	1.6789	0.0196	0.001***
Within-canopy position (*C*)	1	11.6402	0.0716	0.001***	1	11.0292	0.0644	0.001***
L_{BV} x *C*	2	1.1977	0.0147	0.019*	2	0.9442	0.0110	0.156
Surface vs. tissue								
Within-leaf position (L_{ST})	1	6.0058	0.1505	0.001***	1	9.6639	0.2095	0.001***
Within-canopy position (*C*)	5	1.5434	0.1934	0.001***	5	1.3339	0.1446	0.016*
L_{ST} x *C*	5	1.4372	0.1801	0.003**	5	1.1573	0.1255	0.144

Table 5 Effect of within-canopy initial position of European beech (*Fagus sylvatica*) leaves on microbial community composition during leaf decomposition, tested using permutational analyses of variance (PERMANOVA). Dissimilarities among leaves were estimated using the binary version of Jaccard distance. The incubation time (IT), within-canopy initial position (*C*), and their interaction were tested using sample Set#4 (Fig. 1C) with permutations constrained by tree identity (*T*)

	Bacteria				Fungi			
	d.f	*F*-value	R^2	*P*-value	d.f	*F*-value	R^2	*P*-value
Incubation time (IT)	3	2.9060	0.1093	0.001***	3	3.6528	0.1325	0.001***
Within-canopy position (*C*)	1	4.7030	0.0590	0.001***	1	4.2306	0.0511	0.001***
IT x *C*	3	1.4367	0.0541	0.001***	3	1.1752	0.0426	0.026*
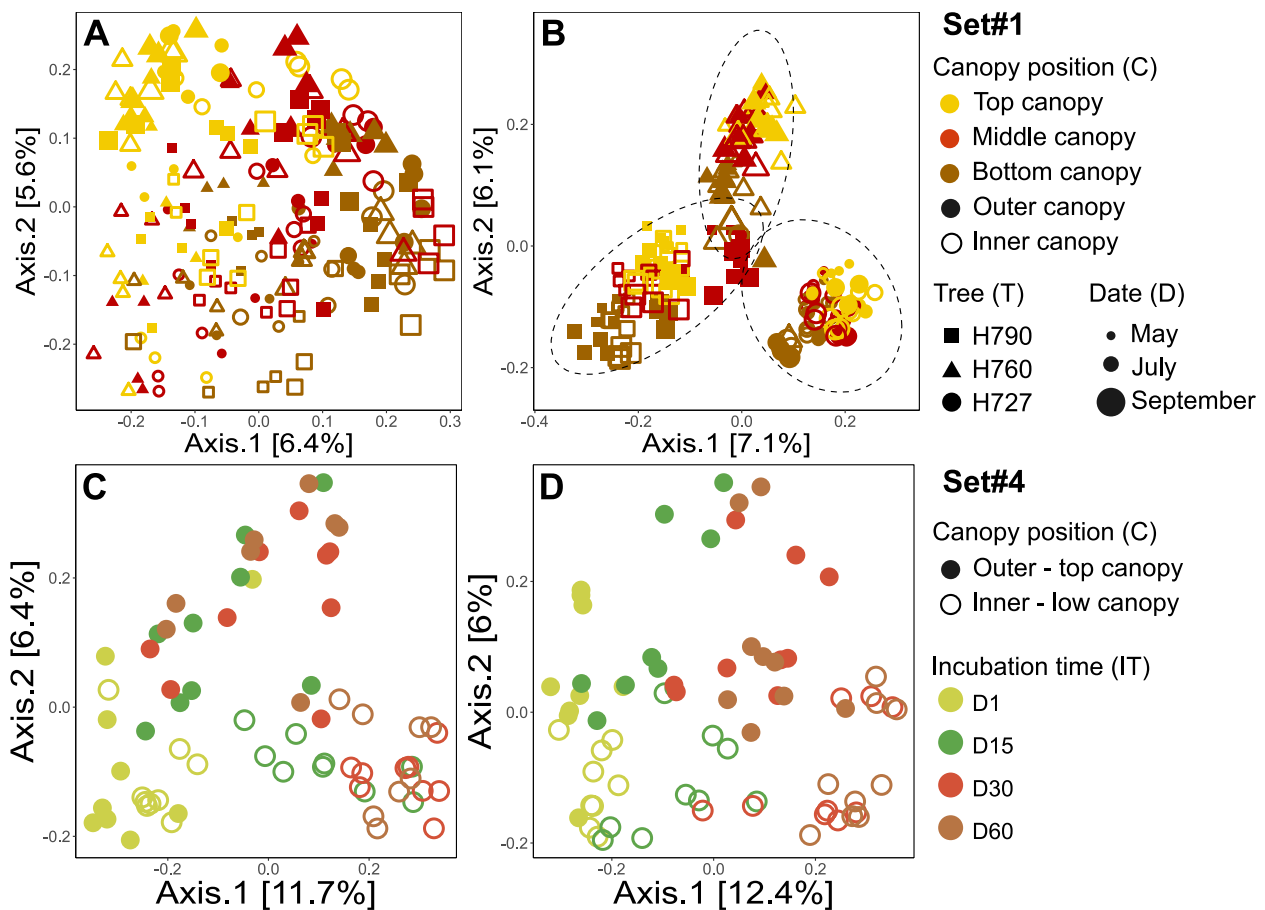
Fig. 2 Variation in leaf microbial community composition in European beech (*Fagus sylvatica*), depending on leaf position within the canopy, tree and date (**A** and **B**) and leaf decomposition stage (**C** and **D**). Dissimilarities among samples were estimated using a binary version of Jaccard distance and represented with PCoA plots. Plots A and C represent bacterial community dissimilarities. Plots B and D represent fungal community dissimilarities. In plots A and B, colors indicate within-canopy vertical and horizontal positions, shapes indicate the tree sampled and size indicate the date of sampling. Confidence ellipses (95%) are drawn around tree (T) for fungal communities. In plots C and D, colors indicate incubation times and shapes indicate leaf initial position within the canopy. All factors had significant effects on microbial composition (Tables 3 and 5 in Appendix)

Code availability

The custom code and/or software application generated during the current study are available in the Harvard Dataverse repository at https://doi.org/10.7910/DVN/FFHAQU.

Authors' contributions

Conceptualization: Corinne Vacher, Lisa Wingate, Tania Fort; methodology: Tania Fort; formal analysis and investigation: Tania Fort; visualization: Tania Fort; writing—original draft preparation: Tania Fort; writing—review and editing: Corinne Vacher; funding acquisition: Corinne Vacher, Lisa Wingate; resources: Emilie Chancerel, Régis Burlett; software: Charlie Pauvert; supervision: Corinne Vacher, Lisa Wingate; project administration: Corinne Vacher. The authors read and approved the final manuscript.

Fort *et al. Annals of Forest Science*   (2023) 80:14

Page 9 of 9

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors gave their informed consent to this publication and its content.

Competing interests
The authors declare no competing interest.

References

Abarenkov K, Nilsson RH, Larsson K-H et al (2018) The UNITE database for molecular identification of fungi – recent updates and future perspectives. New Phytol 219:5. https://doi.org/10.1111/nph.15238

Abarenkov K, Zirk A; Piirmann T et al (2020) UNITE general FASTA release for Fungi. Version 04.02.2020. UNITE Community. https://doi.org/10.15156/BIO/786368

Bahram M, Küngas K, Pent M et al (2022) Vertical stratification of microbial communities in woody plants. Phytobiomes J 6:161–168. https://doi.org/10.1094/pbiomes-06-21-0038-r

Briscoe L, Halperin E, Garud NR (2022) Microbiome source tracking using single nucleotide variants. biorxiv. https://doi.org/10.1101/2022.05.28.493810

Callahan BJ, McMurdie PJ, Rosen MJ et al (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869

Chelius MK, Triplett EW (2001). The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. Microbial ecology 252–263

Cordier T, Robin C, Capdevielle X et al (2012) The composition of phyllosphere fungal assemblages of European beech (*Fagus sylvatica*) varies significantly along an elevation gradient. New Phytol 196:510–519. https://doi.org/10.1111/j.1469-8137.2012.04284.x

Galan M, Razzauti M, Bard E, et al (2016) 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife. mSystems 1:e00032–16. https://doi.org/10.1128/mSystems.00032-16

Jaccard P (1901) Étude comparative de la distribution florale dans une portion des Alpes et du Jura. Bull La Société Vaudoise Des Sci Nat 37:547–579. https://doi.org/10.5169/seals-266450

Laforest-Lapointe I, Paquette A, Messier C, Kembel SW (2017) Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. Nature 546:145–147. https://doi.org/10.1038/nature22399

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet J 17:2803–2809

McKnight DT, Huerlimann R, Bower DS et al (2019) microDecon: a highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. Environ DNA 1:14–25. https://doi.org/10.1002/edn3.11

McMurdie PJ, Holmes S (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. https://doi.org/10.1371/journal.pone.0061217

Miller ET, Svanbäck R, Bohannan BJM (2018) Microbiomes as metacommunities: understanding host-associated microbes through metacommunity ecology. Trends Ecol Evol 33:926–935. https://doi.org/10.1016/j.tree.2018.09.002

Oksanen J, Blanchet G, Friendly M et al (2019) vegan: Community Ecology Package. R package version 2.5–6. https://CRAN.R-project.org/package=vegan

Quast C, Pruesse E, Yilmaz P et al (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res 41:590–596. https://doi.org/10.1093/nar/gks1219

R Development Core Team (2019) R: A language and environment for statistical computing, v.3.5.3. Vienna, Austria: R Foundation for Statistical Computing. [WWW document]. http://www.R-project.org/

Redford AJ, Bowers RM, Knight R et al (2010). The ecology ofthe phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. Environmental microbiology 12:2885–2893

Schoch CL, Seifert KA, Huhndorf S et al (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci 109:6241–6246. https://doi.org/10.1073/pnas.1117018109

Stone BWG, Jackson CR (2019) Canopy position is a stronger determinant of bacterial community composition and diversity than environmental disturbance in the phyllosphere. FEMS Microbiology Ecology 95:fiz032. https://doi.org/10.1093/femsec/fiz032

Terhonen E, Blumenstein K, Kovalchuk A (2019) Forest tree microbiomes and associated fungal endophytes: functional roles and impact on forest health. Forests 10:42. https://doi.org/10.3390/f10010042

Vacher C, Hampe A, Porté AJ et al (2016). The phyllosphere: microbial jungle at the plant-climate interface. Annual Review of Ecology, Evolution, and Systematics 47:1–24

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267. https://doi.org/10.1128/AEM.00062-07

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.