

**Title:** NGS data reveals the phyllosphere microbiome of wheat plants infected by the fungal pathogen *Zymoseptoria tritici*

**Authors:** Didac Barroso-Bergada <sup>1</sup>, Marie Massot <sup>2</sup>, Noémie Vignolles <sup>3</sup>, Julie Faivre d'Arcier <sup>4</sup>, Emilie Chancerel <sup>2, 4</sup>, Erwan Guichoux <sup>2</sup>, Anne-Sophie Walker <sup>3</sup>, Corinne Vacher <sup>4</sup>, David A. Bohan <sup>1</sup>, Valérie Laval <sup>3, \*</sup>, Frédéric Suffert <sup>3, \*, †</sup>

\* égal contribution

† corresponding author: [frederic.suffert@inrae.fr](mailto:frederic.suffert@inrae.fr)

**Addresses:**

<sup>1</sup> UMR Agroécologie, AgroSup Dijon, INRAE, Université de Bourgogne Franche-Comté, 21065 Dijon, France

<sup>2</sup> UMR PGTB, INRAE, Université de Bordeaux, 33610 Cestas, France

<sup>3</sup> Université Paris-Saclay, INRAE, UR BIOGER, 78850 Thiverval-Grignon, France

<sup>4</sup> UMR BIOGECO, INRAE, Université de Bordeaux, 33615 Pessac, France

## Abstract

The fungal pathogen *Zymoseptoria tritici* is the causal agent of Septoria tritici blotch (STB), a major wheat disease in Western Europe. Microorganisms inhabiting wheat leaves might act as beneficial, biocontrol or facilitating agents that could limit or stimulate the development of *Z. tritici*. Improving our understanding of microbial communities in the wheat phyllosphere would lead to new insights into STB management. This resource announcement provides fungal and bacterial metabarcoding datasets obtained by sampling wheat leaves differing in the presence of symptoms caused by *Z. tritici*. Tissues were sampled from three wheat commercial varieties on three sampling dates during a cropping season. Weeds around wheat fields were sampled as well. In total, more than 450 leaf samples were collected. The pathogen *Z. tritici* was quantified using qPCR. We provide the raw metabarcoding datasets, the Amplicon Sequence Variant (ASV) tables obtained after bioinformatic processing, the metadata associated to each sample (sampling date, wheat variety and tissue health condition), a preliminary descriptive analysis of the data, and the code used for bioinformatic and descriptive statistical analysis.

## Key words:

community succession, fungal pathogen, microbial diversity, microbiome, leaf microbiota, phyllosphere, wheat

## 1. Introduction

Wheat crops are exposed to several fungal plant pathogens, including *Zymoseptoria tritici*, the causal agent of Septoria tritici blotch (STB), a major disease in Western Europe

(Fones and Gurr, 2015). In field conditions, wheat leaves host a multitude of other microorganisms — endophytic, epiphytic, pathogenic and saprophytic (Błaszczyk et al., 2021) — some of which interact directly or indirectly with *Z. tritici* (Kerdraon et al., 2019). Several taxa may also have antagonistic or synergistic activity while interacting with other taxa, and could be considered as potential biocontrol agents or facilitating agents that can limit or stimulate STB development (Chaudhry et al., 2020). Maximizing the chance of highlighting important interactions, for instance within co-occurrence network analysis, requires a thorough description of communities, under different biotic and abiotic conditions (Röttjers and Faust, 2018). In this resource announcement, we present fungal and bacterial community datasets collected on wheat leaves over the course of a wheat cropping season, taking into account: (i) the physiologic stage of wheat; (ii) the dynamics of STB development; and (iii) different wheat cultivars. We collected leaf samples in monovarietal plots grown with three wheat cultivars, at three dates over a growing season. One of the wheat varieties carried resistance genes to STB that also potentially impact the development of other constituent taxa of the microbial community. Three types of leaf sections were collected, which differed in the presence of symptoms caused by *Z. tritici*: (i) sections with no STB lesion from a visually healthy leaf; (ii) sections with no STB lesion from a visually symptomatic leaf; and (iii) sections with STB lesions. This dataset could be used to explore the co-occurrence of microbial species and thereby improve our understanding of the community dynamics associated with the development of *Z. tritici* on wheat leaves. Weeds in the margins and edges of cultivated wheat fields can act as alternative hosts for

microbial species present in this crop. For this reason, a second dataset composed of weed leaf samples was collected to get an insight into the host range of the microorganisms associated to wheat leaves and potentially interacting with *Z. tritici* in the agrosystem.

## 2. Methods

### 2.1. Sampling

#### 2.1.1. Wheat

Samples were collected in 2018 at the Grignon experimental station (Yvelines, France; 48° 51' N, 1° 58' E) from three varieties of winter-sown, bread wheat (*Triticum aestivum*). Two varieties, Soissons (SOI) and Apache (APA), were considered susceptible to *Z. tritici* (both rated 5 on the ARVALIS-Institut du Végétal/CTPS scale, from 1 to 9, with 9 corresponding to the most resistant cultivar), while the variety Cellule (CEL), carrying the gene *Stb16q*, was considered to be more resistant (rated 7). Leaf samples of each variety were collected in three plots of 30 m<sup>2</sup>. The three APA and CEL plots were independent experimental plots described in Orellana-Torrejon et al. (2022) while the three SOI plots were delineated within a larger (1 ha) wheat field described in Morais et al. (2016) and Kerdraon et al. (2019). Within each plot, five samples were taken at locations spaced of 1 m apart along a transect. For each sample, three pieces of leaf were collected: an asymptomatic leaf piece taken from a leaf without any STB lesion (G); an asymptomatic leaf piece from a leaf with STB lesion (GS); and a symptomatic leaf piece including a portion of sporulating lesion (S), *i.e.* bearing pycnidia

(*Z. tritici* asexual fruiting bodies). Three sampling campaigns were performed: the first on March 14th (SOI) and 15th (APA and CEL); the second on May 3rd; and the third on June 13th. In March and May, leaf pieces measured 5 cm long. G samples were taken from the central part of the second leaf (F2) of a plant located as close as possible to the sampling point. S and GS samples were collected from the third leaf (F3) of another plant. S samples were taken from the distal part of the leaf and GS samples were cut from the basal part (closer to the stem insertion) of the same leaf. In June, leaf pieces measured 3 cm long because leaves were broader and our goal was to collect an approximately similar amount of tissue on all sampling dates. All leaves were found to be symptomatic in June so we only collected S and GS samples from the third leaf of different plants.

### 2.1.2. Weeds

On July 16th, samples were collected on eight species of weeds to produce a complementary dataset. Some of these weeds presented sparse symptoms, caused by undetermined fungal pathogens that had a very high probability of not being *Z. tritici*, specific to wheat. Five GS and five S samples were collected on *Lolium perenne* (LOLPE) individuals growing within the SOI field and on *Arrhenatherum elatius* (ARREL) individuals growing on a slope 5 m away from the SOI field. These two weed species were dominant weeds at the time of sampling. Five G samples were also collected on *Senecio vulgaris* (SENVU) individuals growing within the SOI field, *Poa annua* (POAAN) individuals growing on a path along the SOI field, *Hordeum murinum*

(HORMU) and *Plantago lanceolata* (PLALA) individuals growing between the field and the path, and *Urtica dioica* (URTDI) and *Geranium molle* (GERMO) individuals growing on the slope 2 m away from the SOI field. All leaf samples were cut with scissors and placed in 2 mL autoclaved collection tubes. They were then brought back to the laboratory and stored at -20°C prior to freeze-drying.

## 2.2. DNA extraction

Total DNA was extracted with the DNeasy Plant Mini kit (Qiagen, France), using a protocol slightly modified from that recommended by Kerdraon et al. (2019). Two autoclaved DNAase-free inox 420C beads were added to each tube and samples were ground at 1500 rpm with the Geno/Grinder® for 30 s, then twice 1 min, with manual shaking between each grinding step. Tubes were then centrifuged for 1 min at 6000 g. Leaf powder and 200 µL of buffer AP1 preheated to 60°C were mixed by vortexing the tubes for 30 s twice at 1500 g, and centrifuging them for 1 min at 3000 g. 250 µL of preheated buffer AP1 and 4.5 µL of RNase A were added to each tube and mixed by vortexing the tubes for 30s twice at 1500 g. After 5 min of rest, 130 µL of buffer P3 was added to each tube, which was then mixed by gentle inversion for 15 s, incubated at -20°C for 10 min and centrifuged for 1 min at 5000 g. The supernatant (450 µL) was transferred to a spin column and centrifuged for 2 min at 20000 g. The filtrate (200 µL) was transferred to a new tube, to which sodium acetate (200 µL, 3M, pH 5) and cold 2-propanol (600 µL) were added. DNA was precipitated by incubation at -20°C for a minimum of 1 h and recovered by centrifugation (20 min, 13000 g). The pellet was washed with cold ethanol (70%),

dried at 50°C for about 30min, and dissolved in 100 µL of AE buffer.

### 2.3. Bacterial 16S amplification

The V5-V6 region of the bacterial 16S rDNA gene was amplified using primers 799F-1115R (Redford et al., 2010; Chelius and Triplett, 2001) to exclude chloroplastic DNA.

To avoid a two-stage PCR protocol and reduce PCR biases, each primer contained the Illumina adaptor sequence, a tag and a heterogeneity spacer, as described in Laforest-

Lapointe et al. (2017) (799F: 5<sup>I</sup>-

CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCC

GATCTxxxxxxxxxxHS-AACMGGATTAGATACCCKG-3<sup>I</sup>; 1115R:5<sup>I</sup>-

AATGATAACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCG

ATCTxxxxxxxxxxHS-AGGGTTGCGCTCGTTG-3<sup>I</sup>, where HS represents a 0-7-base-pair heterogeneity spacer and "x" a 12 nucleotide tag). The PCR mixture (20 µL of final

volume) consisted of 4 µL of buffer Phusion High Fidelity 5X (ThermoFisher) (1X final),

2 µL each of the forward and reverse primers (0.2 µM final), 2 µL of 2 mM dNTPs (200

µM final), 8.2 µL of water, 0.6 µL of SO, DM0, 2 µL of Phusion Hot Start II Polymerase

(ThermoFisher) and 1 µL of DNA template. PCR cycling reactions were conducted on a

Veriti 96-well Thermal Cycler (Applied Biosystems) using the following conditions:

initial denaturation at 98°C for 30s followed by 30 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. Two bacterial strains (*Vibrio*

*splendidus* et *Sulfitobacter pontiacus*) were used as positive controls as they were unlikely

to be found in our samples. The negative PCR controls were represented by PCR mix

without any DNA template. Each PCR plate contained are negative extraction control, three negative PCR controls, one single-strain positive control and one two-strain positive control.

#### **2.4. Fungal ITS amplification**

The ITS1 region of the fungal ITS rDNA gene (Schoch et al., 2012) was amplified using primers ITS1F-ITS2 (White et al., 1990; Gardes and Bruns, 1993). To avoid a two-stage PCR protocol, each primer contained the Illumina adaptor sequence and a tag (ITS1F: 5<sup>1</sup>-CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxxxxxxxxCTTGGTCATTAGAGGAAGTAA-3<sup>1</sup>; ITS2: 5<sup>1</sup>-AATGATAACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTxxxxxxxxxxGCTGCGTTCTTCATCGATGC-3<sup>1</sup>, where "x" is the 12 nucleotide tag). The PCR mixture (20 µL of final volume) consisted of 10 µL of 2X QIAGEN Multiplex PCR Master Mix (2X final), 2 µL each of the forward and reverse primers (0.1 µM final), 4 µL of water, 1 µL of 10 ng/µL BSA and 1 µL of DNA template. 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 94°C for 30 s, 57°C for 90 s, 72°C for 90 s with final extension of 72°C for 10 min. ITS1 amplification was confirmed by electrophoresis on a 2% agarose gel. Two marine fungal strains (*Candida oceanii* and *Yamadazyma barbieri*) were used as positive controls as they were unlikely to be found in our samples. One positive control included 1 µL of 10 ng/µL DNA of *Candida oceanii* only and the other included an equimolar mixture of both

strains. The negative PCR controls were represented by PCR mix without any DNA template. Each PCR plate contained one negative extraction control, three negative PCR controls, one single-strain positive control and one two-strain positive control.

## 2.5. Sequencing

MiSeq sequencing PCR products purification (CleanPCR, MokaScience), library sequencing on an Illumina MiSeq platform (v2 chemistry,  $2 \times 250$  bp) and sequence demultiplexing (with exact index search) were performed at the PGTB sequencing facility (Genome Transcriptome Platform of Bordeaux, Pierroton, France). Fungal ITS1 amplicons were sequenced on three runs and bacterial 16S amplicons were sequenced on four runs.

## 2.6. Bioinformatic treatment

The MiSeq sequences produced were processed using the DADA2 pipeline version 1.22.0 (Callahan et al., 2016) and implemented in R. Primers were identified and removed using CUTADAPT 3.2 (Martin, 2011) and the trimmed sequences were then parsed to the DADA2 algorithm. Chimeras were removed using the removeBimeraDenovo functionality of DADA2. ASVs taxonomic assignment was performed using an implementation of the Naive Bayesian Classifier (Wang et al., 2007) included in the DADA2 pipeline. The databases used for taxonomic assignment were the Silva v138.1 (Quast et al., 2012) and the UNITE all eukaryotes v8.3 (Abarenkov et al., 2021) for 16S and ITS sequences, respectively. Three tables were obtained at the end of this process: an ASV table with the sequence count in each sample; a table with the taxonomic assignment of each ASV

sequence; and a metadata table describing the collection conditions of each sample. The three tables were joined in a phyloseq object using the phyloseq bioconductor package v1.38.0 (McMurdie and Holmes, 2013). To filter out possible contaminants, the combined method of the isContaminant function of the DECONTAM Bioconductor package v1.14.0 (Davis et al., 2018) was used, followed by the decontamination method described in Galan et al. (2016). Moreover, 16S ASVs identified as chloroplastic or mitochondrial with Metaxa2.2.3 (Bengtsson-Palme et al., 2015), or according to their taxonomic assignment in the Silva database, were removed. The remaining ASVs were clustered using the Lulu algorithm (Frøslev et al., 2017) with default parameters. ASVs that could not be assigned to a bacterial or fungal phylum were removed. Finally, ASVs present in less than 1% of the samples were removed to make sure that the data were free of sequencing artifacts and low abundant contaminants (Cao et al., 2021).

## 2.8. Quantification of *Z. tritici* by qPCR

The abundance of *Z. tritici* in wheat tissues was estimated using the quantitative PCR assay developed by Duvivier et al. (2013). The specific set of primers included a forward primer (5'-ATTGGCGAGAGGGATGAAGG-3'), a reverse primer (5'-TTCGTGTCCCAGTGCCTGTA-3'), both leading to an amplification product of 101 pb, and a Taqman fluorogenic probe (5'-ACGACTCGCGGCTTCACCCAACG-3'). The probe was labelled with a FAM fluorescent reporter dye and a BHQ-1 quencher. The quantification reaction was performed with the CFX96 Real time System C1000 Thermal Cycler (BIORAD, USA), using hard shell PCR 96-well WHT/CLR plates. The mix

reaction was composed of reverse and forward primers at 500 nM per reaction, the probe at 500 nM per reaction in a final volume of 25  $\mu$ L, with 5  $\mu$ L of DNA introduced per well. All samples (standard DNA, eDNA to be analyzed, and negative controls) were analyzed with three replicates. The PCR program was 95°C for 10 min and (95°C for 15 s, 60°C for 20 s, 72°C for 40 s) repeated for 40 cycles. The concentration of DNA in the unknown samples was calculated by comparing cycle threshold (Ct) values of the samples with known standard quantities of *Z. tritici* genomic DNA, using a tenfold serial dilution from 5 ng to  $5 \cdot 10^{-3}$  ng per well. Ct values were plotted against the log of the initial concentration of *Z. tritici* genomic DNA to produce the standard curve used for sample quantity determination.

## 2.9. Analysis

Data contained in the phyloseq object were analyzed using the statistic environment R v4.1.2 (R Core Team, 2020) to characterize the fungal and bacterial community composition and to assess the effect of the different experimental factors on these communities. The analysis was performed using only the samples obtained from wheat plants at March and May. Samples obtained in June were not included in the analysis because there were no healthy (G) samples available. ASV counts were transformed using a clr transformation (Aitchison, 1982) to obtain scale-invariant values, avoiding the compositional effect introduced during the sequencing process (Gloor et al., 2017). Then, the phyloseq R package was used to obtain the euclidean distance between samples and to perform a Principal Coordinate Analysis (PCoA). The PCoA was plotted using ggplot2

package v3.3.5 (Wickham, 2016). A permutational multivariate analysis of variance (permanova), performed to assess the effect of the experimental design on the communities, was done using the adonis2 function of the vegan R package v2.5.7(Oksanen et al., 2019) following the experimental formula “tissue  $\times$  date  $\times$  variety/plot”. Alpha diversity measures were obtained using the phyloseq package and fitted in a generalised mixed model using the lme4 R package v1.1-27.1 (Bates et al., 2015). *Z. tritici* qPCR analysis was also fitted in a generalised mixed model using lme4.

### 3. Results

This resource announcement provides two sets of raw sequences files, one set obtained using primers for the fungal ITS region and another obtained using primers for the bacterial 16S region. The sequences are available in the Dataverse files (see section Availability of Data and Materials). The raw and filtered ASV tables obtained during the dereplication and filtering process are provided in the form of phyloseq objects (McMurdie and Holmes, 2013). Each phyloseq object includes the ASV table, a table with the ASV taxonomic assignation and a metadata table. The raw ASV tables also include the positives and negatives control samples used for the filtering. The samples obtained in June, as well as samples obtained from weeds growing in the vicinity of the wheat crop, are included in the phyloseq objects but were not analyzed in the present study. The metadata table includes, for each sample, the wheat variety or weed species sampled, the sampling date, the plot and the wheat cultivar, the visual assessment of symptoms and the *Z. tritici* DNA concentration obtained by qPCR (Figure 1A, Table 1, Suppl. Table S1). The tables

showing the change in number of reads in each sample during the bioinformatic process are also provided in the Dataverse files.

### 3.1. Fungal communities

360 samples were sequenced using ITS primers and gave an average of 31,586 raw fungal sequences per sample with a minimum of 45 reads and a maximum of 246,434 reads per sample. The amplicon sequence variant (ASV) inference process identified an average of 28,178 high quality sequences per sample distributed in 2821 unique ASVs in the 360 samples. The ASV table obtained after the filtering process, which deleted contaminants and low abundant ASVs, was made up of an average of 27,609 sequences per sample distributed between 391 ASVs and 357 samples. 3 samples did not have any sequence after the filtering process. The minimum number of reads in a sample was 20 and the maximum was 223,756. 101 samples of weeds growing close to the field were also sequenced using ITS primers. The bioinformatic process allowed to obtain a mean of 45,631 sequences per sample and a total of 337 ASVs from these weed samples. The minimum number of reads in a sample was 1331 and the maximum was 365,035. The number of reads in each sample at each step of the bioinformatic process is supplied in the Dataverse files.

*Taxonomic composition* — For the wheat dataset, sequences assigned to the Ascomycota represented 74% of the total counts, while sequences assigned to Basidiomycota represented 25% (Figure 2A). As expected, sequences assigned to the genus *Zymoseptoria* were the most abundant (60% of the sequences). *Zymoseptoria* was also more abundant in symptomatic than in asymptomatic leaf samples. It was also slightly more abundant in

the Soissons (SOI) and the Apache (APA) varieties, than in the Cellule (CEL) cultivar, which is less susceptible because it carries the *Stb16q* resistance gene.

*Alpha diversity* — Fungal community richness, evenness and diversity differed significantly between dates and tissue health conditions (Suppl. Figure S1, Suppl. Table S2). Wheat variety had minor effect in all alpha diversity measures, being significant only for richness and diversity.

*Beta diversity* — The composition of wheat foliar fungal communities differed significantly among dates, varieties and tissue health conditions (Figure 1B, Suppl. Figure S1, Table 1, Suppl. Table S2). Tissue was the most important factor, explaining 10% of the variance. Samples collected in June were not included in the permutational analysis of variance to avoid a potential bias caused by the absence of healthy leaves (G samples) at that time.

### 3.2. Bacterial communities

The 360 samples used for ITS sequencing were also sequenced using 16S primers, obtaining an average of 40,724 raw bacterial sequences per sample with a minimum of 0 read and a maximum of 92520 reads per sample. The amplicon sequence variant (ASV) inference process identified a mean of 31,969 high quality sequences shared between 12,349 unique ASVs in 350 samples. 10 samples did not generate enough sequences to perform the ASV inference. The ASV table obtained after the filtering process, carried out to delete contaminants and low abundant ASVs, was composed of an average of 13,964 sequences per sample distributed between 1495 ASVs and 340 samples. The minimum number of

reads in a sample was 2 and the maximum was 71,051. 102 samples of weeds growing surrounding the wheat plots were also sequenced using 16S primers. The bioinformatic process produced an average of 29,991 sequences per sample and 1,068 unique ASVs from the weed samples. The minimum number of reads in a sample was 30 and the maximum was 70,999. The number of reads of each sample at each step of the bioinformatic process is supplied in the Dataverse files.

*Taxonomic composition* — The most abundant bacterial phyla were Proteobacteria (39% of sequence counts), followed by Actinobacteria (35%), Bacteroidetes (12%) and Firmicutes (11%) (Figure 2B). Proteobacteria were more evident in later sampling dates while Actinobacteria were more present at the March sampling. Tissue condition and wheat variety did not seem to have an important effect on the community composition of the phyla.

*Alpha diversity* — Bacterial community richness, evenness and diversity differed significantly between tissue health conditions (Suppl. Figure S1, Suppl. Table S3). On the other side, sampling date had effect on richness and evenness measures, while wheat variety only had effect on the diversity measure.

*Beta diversity* — The composition of the bacterial communities of wheat leaves differed significantly among dates, varieties and tissue conditions (Figure 1C, Suppl. Figure S1, Table 1, Suppl. Table S3). As with the fungal component of the community, date was the most important structuring factor, explaining some 8% of the variance in the bacterial community (Table 1). Samples obtained in June were not included in the permutational analysis to avoid a potential for bias caused by the absence of healthy leave

samples on that sample date.

#### 4. Conclusions

Preliminary statistical analyses revealed that sampling date, wheat variety and STB symptoms had significant effects on fungal and bacterial communities of the wheat phyllosphere. While the three factors tested structured the community, the date of sampling exhibited the strongest effect. As expected, we found congruence between the presence of *Z. tritici*, assessed by eye (STB symptoms on the leaves) and by qPCR (concentration of *Z. tritici* DNA within the leaf tissues). This finding confirms the relevance of the sampling strategy to meet the assigned objectives. Co-occurrence network analyses could be used to characterize the dynamics of the community associated with *Z. tritici* and might help identify individual taxa of interest as potential biocontrol or beneficial agents to improve "wheat health". The weed dataset could also be examined for *Z. tritici* interactions with different communities present on non-crop plants within and in the margins of the field.

#### Availability of Data and Materials

The sequence datasets were deposited in NCBI SRA in bioproject PRJNA803042 (<https://www.ncbi.nlm.nih.gov/bioproject/803042>). The biosample accession numbers are SAMN25610777 to SAMN25611238. Bioinformatic scripts and raw and filtered ASV tables in R phyloseq format were deposited in Dataverse (<https://doi.org/10.15454/QTXFP9>). The tables showing variation in sequence counts during

the bioinformatic process and the scripts used for data processing and statistical analysis were included in the Dataverse deposit.

### **Author contributions**

Author contributions were as follows: (i) D.A.B., C.V., F.S. and V.L coordinated the study; (ii) F.S., V.L., A.-S.W. and M.M. designed the sampling, collected samples and metadata; (iii) N.V., J.F.A., E.C. and E.G. developed protocols and performed the molecular biology work; (iv) D.B.B. performed data analysis and wrote the manuscript; (v) all authors edited the manuscript.

### **Acknowledgments**

We thank all members of the *Consortium Biocontrôle* for their support for the project. We thank Céline Lalanne, Adline Delcamp, Christophe Boury and all other members of PGTB (Genome Transcriptome Platform of Bordeaux) for their support in molecular biology, and Gaétan Burgaud and Frédéric Garabetian for kindly providing marine strains used as positive controls. We also thank Frédéric Barraquand, Stéphane Robin, Charlie Pauvert and Andreas Makiola for helpful discussions at the beginning of the project. We acknowledge support from the ANR NGB project (ANR-17-CE32-0011). PTGB and BIOGER received support from *Investissements d'avenir* and *convention attributive d'aide EquipEx Xyloforest* (ANR-10-EQPX-16-01) and from *Saclay Plant Sciences-SPS* (ANR-17-EUR-0007), respectively.

## References

K. Abarenkov, A. Zirk, T. Piirmann, R. Pöhönen, F. Ivanov, R. H. Nilsson, and U. Kõljalg. UNITE general FASTA release for eukaryotes 2, may 2021, UNITE Community.

J. Aitchison. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, Jan 1982. ISSN 0035-9246. doi: 10.1111/j.2517-6161.1982.tb01195.x.

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.

J. Bengtsson-Palme, M. Hartmann, K. M. Eriksson, C. Pal, K. Thorell, D. G. J. Larsson, and R. H. Nilsson. METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol. Ecol. Resour.*, 15(6):1403–1414, Nov 2015. ISSN 1755-098X. doi: 10.1111/1755-0998.12399.

L. Błaszczyk, S. Salamon, and K. Mikołajczak. Fungi inhabiting the wheat endosphere. *Pathogens*, 10(10), 2021. ISSN 2076-0817. doi: 10.3390/pathogens10101288. URL <https://www.mdpi.com/2076-0817/10/10/1288>.

B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*, 13:581–583, Jul 2016. ISSN 1548-7105. doi: 10.1038/nmeth.3869.

Q. Cao, X. Sun, K. Rajesh, N. Chalasani, K. Gelow, B. Katz, V. H. Shah, A. J. Sanyal, and E. Smirnova. Effects of rare microbiome taxa filtering on statistical analysis. *Front. Microbiol.*, 2021. ISSN 1664-302X. doi: 10.3389/fmicb.2020.607325.

V. Chaudhry, P. Runge, P. Sengupta, G. Doeblemann, J. E. Parker, and E. Ke-men.

Shaping the leaf microbiota: plant–microbe–microbe interactions. *Journal of Experimental Botany*, 72(1):36–56, 09 2020. ISSN 0022-0957. doi: 10.1093/jxb/eraa417. URL <https://doi.org/10.1093/jxb/eraa417>.

M. K. Chelius and E. W. Triplett. The diversity of archaea and bacteria in association with the roots of *Zea mays*. *Microbial Ecology*, 41(3):252–263, 2001. ISSN 00953628, 1432184X. URL <http://www.jstor.org/stable/4251818>.

N. M. Davis, D. M. Proctor, S. P. Holmes, D. A. Relman, and B. J. Calla-han. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, 6(1):1–14, Dec 2018. ISSN 2049-2618. doi: 10.1186/s40168-018-0605-2.

M. Duvivier, G. Dedeurwaerder, M. De Proft, J.-M. Moreau, and A. Legrèvre. Real-time PCR quantification and spatio-temporal distribution of airborne inoculum of *Mycosphaerella graminicola* in Belgium. *Eur. J. Plant Pathol.*, 137(2):325–341, Oct 2013. ISSN 1573-8469. doi: 10.1007/s10658-013-0245-0.

H. Fones and S. Gurr. The impact of *Septoria tritici* blotch disease on wheat: An EU perspective. *Fungal Genet. Biol.*, 79:3–7, Jun 2015. ISSN 1087-1845. doi: 10.1016/j.fgb.2015.04.004.

T. G. Frøslev, R. Kjøller, H. H. Bruun, R. Ejrnæs, A. K. Brunbjerg, C. Pietroni, and A. J. Hansen. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat. Commun.*, 8(1188):1–11, Oct 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01312-x.

M. Galan, M. Razzauti, E. Bard, M. Bernard, C. Brouat, N. Charbonnel, A. Dehne-Garcia, A. Loiseau, C. Tatard, L. Tamisier, M. Vayssier-Taussat, H. Vignes, and J.-F. Cosson. 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in Wildlife. *mSystems*, 1(4), Jul 2016. doi: 10.1128/mSystems.00032-16.

M. Gardes and T. D. Bruns. ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. *Mol. Ecol.*, 2(2):113–118, Apr 1993. ISSN 0962-1083. doi: 10.1111/j.1365-294x.1993.tb00005.x.

G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, and J. J. Egozcue. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.*, 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.02224.

L. Kerdraon, M. Barret, V. Laval, and F. Suffert. Differential dynamics of microbial community networks help identify microorganisms interacting with residue-borne pathogens: the case of *Zymoseptoria tritici* in wheat. *Microbiome*, 7(1):1–17, Dec 2019. ISSN 2049-2618. doi: 10.1186/s40168-019-0736-0.

I. Laforest-Lapointe, A. Paquette, C. Messier, and S. W. Kembel. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature*, 546:145–147, Jun 2017. ISSN 1476-4687. doi: 10.1038/nature22399.

M. Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12, May 2011. ISSN 2226-6089. doi: 10.14806/ej.17.1.200.

P. J. McMurdie and S. Holmes. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4):e61217, Apr 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0061217.

D. Morais, S. Gélisse, V. Laval, I. Sache, and F. Suffert. Inferring the origin of primary inoculum of *Zymoseptoria tritici* from differential adaptation of resident and immigrant populations to wheat cultivars. *Eur. J. Plant Pathol.*, 145(2):393–404, Jun 2016. ISSN 1573-8469. doi: 10.1007/s10658-015-0853-y.

J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O’Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. *vegan: Community Ecology Package*, 2019. URL <https://CRAN.R-project.org/package=vegan>. R package version 2.5-6.

C. Orellana-Torrejon, T. Vidal, A.-L. Boixel, S. Gélisse, S. Saint-Jean, and F. Suffert. annual dynamics of *Zymoseptoria tritici* populations in wheat cultivar mixtures: a compromise between the efficiency and durability of a recently broken-down resistance gene? *Plant Pathol.*, 71:289–303, Aug 2021. ISSN 0032-0862. doi: 10.1111/ppa.13458.

C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1): D590–D596, 11 2012. ISSN 0305-1048. doi: 10.1093/nar/gks1219. URL <https://doi.org/10.1093/nar/gks1219>.

R Core Team. *R: A Language and Environment for Statistical Computing*. R foundation for statistical computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.

A. J. Redford, R. M. Bowers, R. Knight, Y. Linhart, and N. Fierer. The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environmental Microbiology*, 12(11):2885–2893, 2010. doi: 10.1111/j.1462-

2920.2010.02258.x

L. Röttjers and K. Faust. From hairballs to hypotheses-biological insights from microbial networks. *FEMS Microbiol. Rev.*, 42(6):761–780, Nov 2018. ISSN 1574-6976. doi: 10.1093/femsre/fuy030.

C. L. Schoch, K. A. Seifert, S. Huhndorf, V. Robert, J. L. Spouge, C. A. Levesque, W. Chen, and F. B. Consortium. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences*, 109(16):6241– 6246, 2012. ISSN 0027-8424. doi: 10.1073/pnas.1117018109. URL <https://www.pnas.org/content/109/16/6241>.

Q. Wang, G. M. Garrity, J. M. Tiedje, and J. R. Cole. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.*, 73(16):5261–5267, Aug 2007. ISSN 0099-2240. doi: 10.1128/AEM.00062-07.

T. J. White, T. D. Bruns, S. B. Lee, J. W. Taylor, and J. J. Sninsky. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: a Guide to Methods and Applications*, 31:315–322, Jan 1990. URL [https://www.researchgate.net/publication/262687766\\_Amplification\\_and\\_Direct-Sequencing\\_of\\_Fungal\\_Ribosomal\\_RNA\\_Genes\\_for\\_Phylogenetics](https://www.researchgate.net/publication/262687766_Amplification_and_Direct-Sequencing_of_Fungal_Ribosomal_RNA_Genes_for_Phylogenetics).

H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL <https://ggplot2.tidyverse.org>.

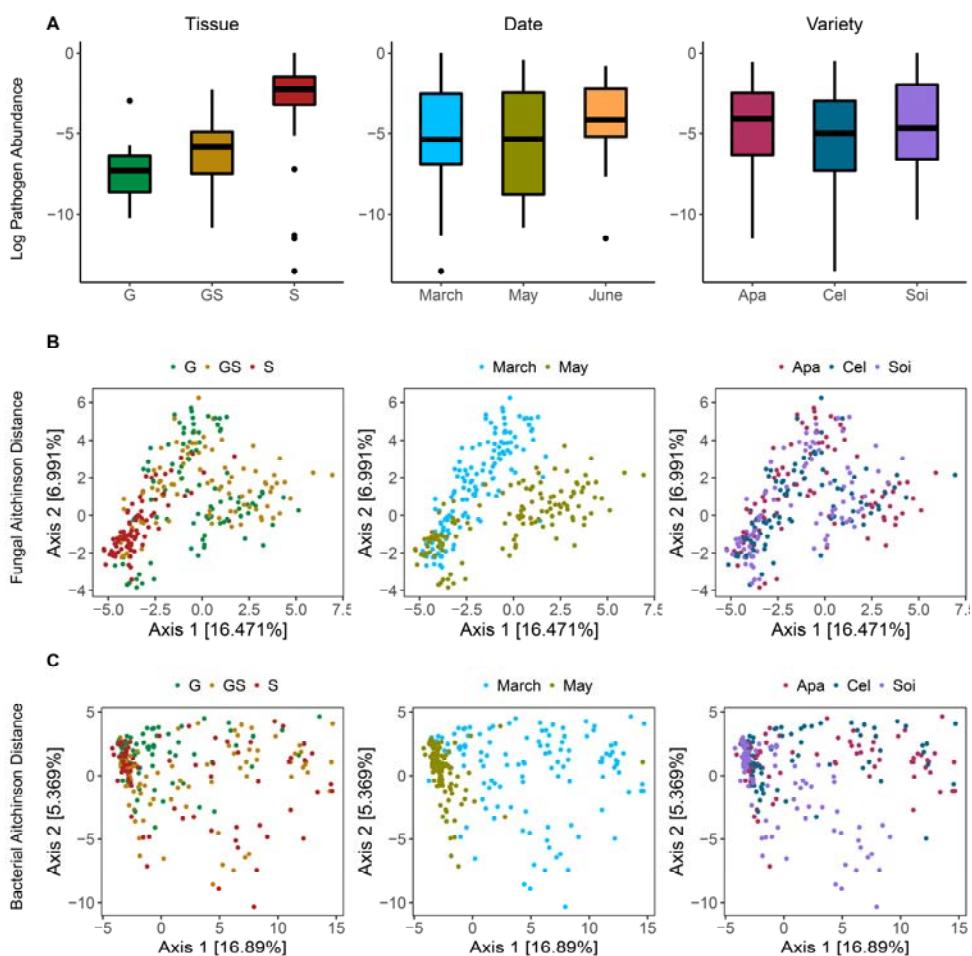
**Table 1. Permutational multivariate analysis of variance between samples.** The Aitchison distance (Aitchison, 1982) between samples in the abundance matrix is used as a distance metric. The analysis was performed for the fungal and bacterial ASV tables separately using the R vegan package (Oksanen et al., 2019). Factors were: Tissue, corresponding to *Septoria tritici* blotch symptomatology; Date, corresponding to the sampling dates; and Variety, corresponding to the wheat varieties sampled. The plots were considered nested to the wheat variety. Values of  $p =$  \*  $< .05$ ; \*\*  $< .01$ ; \*\*\*  $< .001$ .

<b>Fungi</b>	<b>Df</b>	<b>Sum of squares</b>	<b>R2</b>	<b>F</b>	<b>P value</b>
Tissue	2	1369.149	0.051	12.319	***
Date	2	2856.615	0.107	25.702	***
Variety	2	615.13	0.023	5.534	***
Tissue:Date	3	968.617	0.036	5.81	***
Tissue:Variety	4	370.541	0.014	1.667	***
Date:Variety	4	803.509	0.03	3.615	***
Tissue:Date:Variety	6	488.472	0.018	1.465	***
Tissue:Date:Variety:Plot	48	3510.502	0.131	1.316	***
Residual	285	15838.138	0.591		
Total Bacterial	356	26820.673	1		

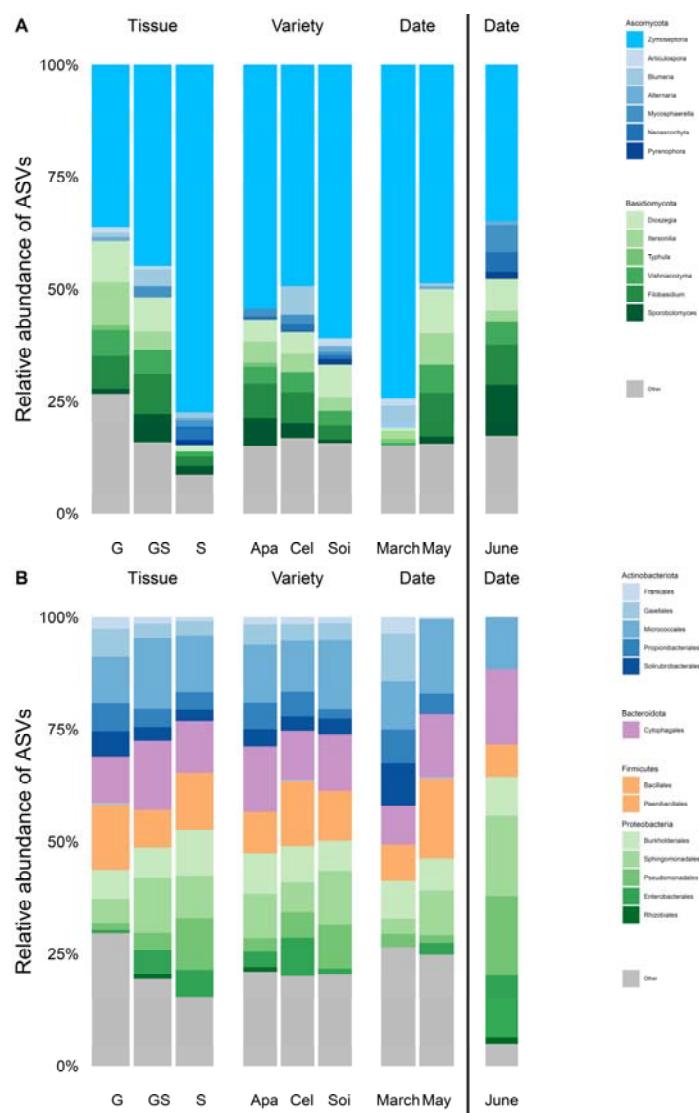
  

<b>Bacteria</b>	<b>Df</b>	<b>Sum of squares</b>	<b>R2</b>	<b>F</b>	<b>P value</b>
Tissue	2	822.805	0.017	3.699	***
Date	2	6153.514	0.126	27.664	***
Variety	2	1189.58	0.024	5.348	***
Tissue:Date	3	856.926	0.017	2.568	***
Tissue:Variety	4	574.873	0.012	1.292	*
Date:Variety	4	1918.814	0.039	4.313	***
Tissue:Date:Variety	6	909.533	0.019	1.363	**
Tissue:Date:Variety:Plot	48	6775.259	0.138	1.269	***
Residual	268	29806.437	0.608		
Total	339	49007.741	1		

**Figure 1. Community composition of wheat leaves.** **A.** Abundance of the pathogen (*Zymoseptoria tritici*) on wheat leaves measured using qPCR. Community composition is represented as the relative abundance of each taxonomic unit for each experimental factor, at the genus level for fungi communities and order level for bacterial. **B, C.** Principal Coordinates Analysis plots showing the similarity of fungal (B) and bacterial (C) communities from different samples. The ordination was performed using the Aitchinson distance (Aitchison, 1982). Plots were colored by: the *Septoria tritici* blotch symptoms, G corresponding to leaf samples collected on asymptomatic leaves, GS corresponding to green parts of a symptomatic leaf, and S to symptomatic parts of a leaf; the sampling seasons, March and May; and the wheat varieties sampled, Apache (APA), Cellule (CEL), and Soissons (SOI).



**Figure 2. Barplots showing the relative abundance of different taxons along different dates, wheat varieties and leaves tissue conditions.** Barplot **A** shows the relative abundance of different fungal genus. Barplot **B** shows the relative abundance of different bacterial orders. Barplots are separated by: Septoria leaf blotch symptoms, G corresponding to leaf samples collected on asymptomatic leaves, GS corresponding to green parts of a symptomatic leaf and S to symptomatic part of a leaf; the sampling seasons, March, May and June (June is separated by a bar because no asymptomatic samples (G) where available).



**Suppl. Table S1. Analysis of Variance (ANOVA) of the fitted linear mixed model of the abundance of *Zymoseptoria tritici*.** *Z. tritici* abundance was obtained by qPCR. Values of p = \* < .05; \*\* < .01; \*\*\* < .001.

	Chisq	Df	Pr..Chisq.
Tissue	201.608	2	***
Variety	18.065	2	***
Date	13.743	2	**
Tissue:Variety	37.47	4	***
Tissue:Date	20.069	3	***
Variety:Date	30.747	4	***
Tissue:Variety:Date	35.217	6	***

**Suppl. Table S2. General Mixed models of three alpha diversity fungal community measures.** Observed richness, Shannon diversity index and inverse Simpson evenness index are modelled. Richness is modeled using a Negative Binomial distribution while Diversity and Evenness are modelled using a Gaussian distribution. Fixed factors are: Septoria leaf blotch symptoms, G corresponding to green parts of asymptomatic leaves, GS corresponding to green parts of symptomatic leaves and S to symptomatic part of symptomatic leaves; the sampling seasons, March and May; and the wheat varieties sampled, Apache (APA), Cellule (CEL), and Soissons (SOI). Plot is considered a random factor. Values of  $p = * < .05$ ;  $** < .01$ ;  $*** < .001$ .

Richness				
Fixed-effects				
	Estimate	Std-error	z	Pval
(Intercept)	-5.837	0.144	-40.611	***
VarietyCel	0.222	0.133	1.676	.
VarietySoi	0.306	0.136	2.247	*
DateMay	0.417	0.113	3.705	***
TissueGS	-0.825	0.127	-6.476	***
TissueS	-2.338	0.14	-16.742	***
Random-effects				
Group	Name	Variance	Std-Dev	
Plot	Intercept	0.003	0.052	
Diversity				
Fixed-effects				
	Estimate	Std-error	z	Pval
(Intercept)	-10.324	0.487	-21.182	***
VarietyCel	0.791	0.573	1.38	.
VarietySoi	1.25	0.573	2.18	*
DateMay	2.156	0.273	7.887	***
TissueGS	-1.971	0.206	-9.557	***
TissueS	-3.217	0.595	-5.406	***
Random-effects				
Group	Name	Variance	Std-Dev	
Plot	Intercept	6.948	2.636	
Evenness				
Fixed-effects				
	Estimate	Std-error	z	Pval
(Intercept)	-11.708	0.408	-28.669	***
VarietyCel	0.388	0.534	0.726	.
VarietySoi	0.78	0.545	1.43	.
DateMay	1.063	0.183	5.802	***
TissueGS	-1.258	0.202	-6.217	***
TissueS	-2.8	0.629	-4.453	***
Random-effects				
Group	Name	Variance	Std-Dev	
Plot	Intercept	0.117	0.342	

**Suppl. Table S3. General Mixed models of three alpha diversity bacterial community measures.** Observed richness, Shannon diversity index and inverse Simpson evenness index are modelled. Richness is modeled using a Negative Binomial distribution while Diversity and Evenness are modelled using a Gaussian distribution. Fixed factors are: Septoria leaf blotch symptoms, G corresponding to green parts of asymptomatic leaves, GS corresponding to green parts of symptomatic leaves and S to symptomatic part of symptomatic leaves; the sampling seasons, March and May; and the wheat varieties sampled, Apache (APA), Cellule (CEL), and Soissons (SOI). Plot is considered a random factor. Values of  $p = * < .05$ ;  $** < .01$ ;  $*** < .001$ .

Richness				
Fixed-effects				
	Estimate	Std-error	z	Pval
(Intercept)	-3.621	0.113	-31.979	***
VarietyCel	0.243	0.134	1.812	.
VarietySoi	0.098	0.13	0.75	
DateMay	0.16	0.078	2.059	*
TissueGS	-0.66	0.093	-7.118	***
TissueS	-0.673	0.097	-6.936	***
Random-effects				
Group	Name	Variance	Std-Dev	
Plot	Intercept	0.013	0.113	
Diversity				
Fixed-effects				
	Estimate	Std-error	z	Pval
(Intercept)	-5.376	0.077	-69.799	***
VarietyCel	0.014	0.092	0.152	
VarietySoi	-0.532	0.164	-3.239	**
DateMay	-0.219	0.163	-1.348	
TissueGS	-0.462	0.101	-4.587	***
TissueS	-1.116	0.139	-8.011	***
Random-effects				
Group	Name	Variance	Std-Dev	
Plot	Intercept	0	0	
Evenness				
Fixed-effects				
	Estimate	Std-error	z	Pval
(Intercept)	-9.634	0.203	-47.413	***
VarietyCel	0.073	0.269	0.272	
VarietySoi	0.398	0.271	1.47	
DateMay	0.946	0.134	7.035	***
TissueGS	-1.051	0.142	-7.384	***
TissueS	-1.567	0.168	-9.32	***
Random-effects				
Group	Name	Variance	Std-Dev	
Plot	Intercept	0.036	0.191	

**Supplementary figure S1. Boxplot showing three alpha diversity community measures for fungal and bacterial data.** Observed richness, Shannon diversity index and inverse Simpson evenness index are plotted. Boxplots are separated by: Septoria tritici blotch symptoms, G corresponding to green parts of asymptomatic leaves, GS corresponding to green parts of symptomatic leaves and S to symptomatic part of symptomatic leaves; the sampling seasons, March and May; and the wheat varieties sampled, Apache (APA), Cellule (CEL), and Soissons (SOI). Plot was treated as a random factor.

