



**HAL**  
open science

## Metagenomic next generation sequencing (mNGS) data reveals the phyllosphere microbiome of wheat plants infected by the fungal pathogen *Zymoseptoria tritici*

Didac Barroso-Bergadà, Marie Massot, Noémie Vignolles, Julie Faivre d’Arcier, Emilie Chancerel, Erwan Guichoux, Anne-Sophie Walker, Corinne Vacher, David Bohan, Valérie Laval, et al.

### ► To cite this version:

Didac Barroso-Bergadà, Marie Massot, Noémie Vignolles, Julie Faivre d’Arcier, Emilie Chancerel, et al.. Metagenomic next generation sequencing (mNGS) data reveals the phyllosphere microbiome of wheat plants infected by the fungal pathogen *Zymoseptoria tritici*. *Phytobiomes Journal*, 2022, 7 (2), pp.281-287. 10.1094/PBIOMES-02-22-0008-FI. hal-03689876

**HAL Id: hal-03689876**

**<https://hal.science/hal-03689876>**

Submitted on 18 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



## RESEARCH

# Metagenomic Next-Generation Sequencing (mNGS) Data Reveal the Phyllosphere Microbiome of Wheat Plants Infected by the Fungal Pathogen *Zymoseptoria tritici*

Didac Barroso-Bergadà,<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> and Frédéric Suffert<sup>3,†</sup>

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

<sup>2</sup> UMR Genome Transcriptome Platform of Bordeaux (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

Accepted for publication 11 May 2022.

## 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 article provides fungal and bacterial metabarcoding datasets obtained by sampling wheat leaves with and without symptoms caused by *Z. tritici*. Tissues were sampled from three commercial wheat 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 quantitative PCR. We provide the raw metabarcoding datasets, the amplicon sequence variant tables obtained after bioinformatic processing, the metadata associated with 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.

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

†Corresponding author: F. Suffert; frederic.suffert@inrae.fr

V. Laval and F. Suffert contributed equally to this work.

**Author contributions:** D.A.B., C.V., F.S., and V.L. coordinated the study; F.S., V.L., A.-S.W., and M.M. designed the sampling and collected samples and metadata; N.V., J.F.A., E.C., and E.G. developed protocols and performed the molecular biology work; D.B.B. performed data analysis and wrote the manuscript; and all authors edited the manuscript.

**Funding:** Support was provided by the Agence Nationale de la Recherche (ANR) NGB project (ANR-17-CE32-0011), Investissements d'avenir and convention attributive d'aide EquipEx Xyloforest (ANR-10-EQPX-16-01) to PGTB, and Saclay Plant Sciences-SPS (ANR-17-EUR-0007) to BIOGER.

**e-Xtra:** Supplementary material is available online.

The author(s) declare no conflict of interest.

Copyright © 2022 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.



Wheat crops are exposed to many 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 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 cooccurrence network analysis) requires a thorough description of communities under different biotic and abiotic conditions (Röttjers and Faust 2018). In this article, we present fungal and bacterial community datasets collected on wheat leaves over the course of a wheat cropping season, taking into account (i) the physiological stage of wheat, (ii) the dynamics of STB development, and (iii) different wheat

cultivars. We collected leaf samples in monovarietal plots of 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 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 cooccurrence 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 the 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 with wheat leaves and potentially interacting with *Z. tritici* in the agroecosystem.

## MATERIALS AND METHODS

**Sampling.** *Wheat.* Samples were collected in 2018 at the Grignon experimental station (Yvelines, France) 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 of 1 to 9, with 9 being 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 by Orellana-Torres et al. (2022) while the three SOI plots were delineated within a larger (1 ha) wheat field described by Morais et al. (2016) and Kerdraon et al. (2019). Within each plot, five samples were taken at locations spaced 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 lesions (G), an asymptomatic leaf piece from a leaf with STB lesions (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 14 March (SOI) and 15 March (APA and CEL), the second on 3 May, and the third on 13 June 13. 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; therefore, we only collected S and GS samples from the third leaf of different plants.

*Weeds.* On 16 July, 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*. Five GS and five S samples were collected from *Lolium perenne* individuals growing within the SOI field and on *Arrhenatherum elatius* individuals growing from 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 from *Senecio vulgaris* individuals growing within the SOI field, *Poa annua* individuals growing on a path by the SOI field, *Hordeum murinum* and

*Plantago lanceolata* individuals growing between the field and the path, and *Urtica dioica* and *Geranium molle* 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.

**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 1,500 rpm with the Geno/Grinder for 30 s, then twice for 1 min, with manual shaking between each grinding step. Tubes were then centrifuged for 1 min at 6,000 × g. Leaf powder and 200 µl of buffer AP1 preheated to 60°C were mixed by vortexing the tubes for 30 s twice at 1,500 × g, and centrifuging them for 1 min at 3,000 × g. To each tube, 250 µl of preheated buffer AP1 and 4.5 µl of RNase A were added and mixed by vortexing the tubes for 30 s twice at 1,500 × 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 5,000 × g. The supernatant (450 µl) was transferred to a spin column and centrifuged for 2 min at 20,000 × g. In a new tube, 200 µl of the filtrate was transferred, with 200 µl of sodium acetate (3 M, pH 5) and 600 µl of cold isopropanol. DNA was precipitated by incubation at -20°C for a minimum of 1 h and recovered by centrifugation (20 min, 13,000 × g). The pellet was washed with cold ethanol (70%), dried at 50°C for approximately 30 min, and dissolved in 100 µl of AE buffer.

**Bacterial 16S amplification.** The V5-V6 region of the bacterial 16S ribosomal DNA (rDNA) gene was amplified using primers 799F and 1115R (Chelius and Triplett 2001; Redford et al. 2010) 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 by Laforest-Lapointe et al. (2017) (799F: 5'-CAAGCA GAAGACGGC ATACGAGATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTxxxxxxxxxxxxHS-AACMGGATTAGATAC CCKG-3' and 1115R: 5'-AATGATACGGCGACCACCGAGATCT AACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxxxxxxxx HS-AGGGTTGCGCTCGTTG-3', where HS represents a 0-to-7-bp heterogeneity spacer and x represents a 12-nucleotide tag). The PCR mixture (20 µl of final volume) consisted of 4 µl of buffer Phusion High Fidelity 5× (Thermo Fisher) (1× 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 dimethyl sulfoxide, 2 µl of Phusion Hot Start II Polymerase (Thermo Fisher), 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 30 s; followed by 30 cycles at 98°C for 15 s, 60°C for 30 s, and 72°C for 30 s; with a final extension of 72°C for 10 min. Two bacterial strains (*Vibrio splendidus* and *Sulfitobacter pontiacus*) were used as positive controls because 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 a negative extraction control, three negative PCR controls, one single-strain positive control, and one two-strain positive control.

**Fungal internal transcribed spacer amplification.** The internal transcribed spacer 1 (ITS1) region of the fungal ITS rDNA gene (Schoch et al. 2012) was amplified using primers ITS1F and ITS2 (Gardes and Bruns 1993; White et al. 1990). To avoid a two-stage PCR protocol, each primer contained the Illumina adaptor sequence and a tag (ITS1F: 5'-CAAGCAGAAGACGGCATA CGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

xxxxxxxxxxxCTTGGTCATTTAGAGGAAGTAA-3' and ITS2: 5'-AATGATACGGCGACCACCGAGATCTACTCTTCCCT ACACGACGCTCTCCGATCTxxxxxxxxxxxGCTGCGTTCCT CATCGATGC-3', where x is the 12-nucleotide tag). The PCR mixture (20  $\mu$ l of final volume) consisted of 10  $\mu$ l of 2 $\times$  Qiagen Multi-plex PCR Master Mix (2 $\times$  final), 2  $\mu$ l each of the forward and reverse primers (0.1  $\mu$ M final), 4  $\mu$ l of water, 1  $\mu$ l of Bovine Serum Albumen (BSA) at 10 ng/ $\mu$ l, and 1  $\mu$ l 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, and 72°C for 90 s; with a final extension of 72°C for 10 min. ITS1 amplification was confirmed by electrophoresis on a 2% agarose gel. Two marine fungal species (*Candida oceanii* and *Yamadazyma barbieri*) were used as positive controls because they were unlikely to be found in our samples. One positive control included 1  $\mu$ l of DNA of *C. oceanii* only at 10 ng/ $\mu$ l and the other included an equimolar mixture of both species. 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-species positive control, and one two-species positive control.

**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 Genome Transcriptome Platform of Bordeaux sequencing facility (PGTB, Pierroton, France). Fungal ITS1 amplicons were sequenced on three runs and bacterial 16S amplicons were sequenced on four runs.

**Bioinformatic treatment.** The MiSeq sequences produced were processed using the DADA2 pipeline version 1.22.0 (Callahan et al. 2016) 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. Taxonomic assignment of amplicon sequence variants (ASVs) 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 by 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 <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).

**Quantification of *Z. tritici* by quantitative PCR.** The abundance of *Z. tritici* in wheat tissues was estimated using the quantitative PCR (qPCR) assay developed by Duvivier et al. (2013). The specific set of primers included a forward primer (5'-ATTGGCGA GAGGGATGAAGG-3') and a reverse primer (5'-TTCGTGTCCC

AGTGCGTGTA-3'), both leading to an amplification product of 101 bp, and a TaqMan fluorogenic probe (5'-ACGACTCGCGGCT TCCACCAACG-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 (Bio-Rad), using hard-shell PCR 96-well WHT/CLR plates. The mix reaction was composed of reverse and forward primers at 500 nM/reaction and the probe at 500 nM/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 followed by 95°C for 15 s, 60°C for 20 s, and 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 10-fold serial dilution from 5 to 5  $\times$  10<sup>-3</sup> ng per well. Ct values were plotted against the log<sub>10</sub> of the initial concentration of *Z. tritici* genomic DNA to produce the standard curve used for sample quantity determination.

**Analysis.** Data contained in the phyloseq object were analyzed using the statistical 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 in March and May. Samples obtained in June were not included in the analysis because there were no healthy (G) samples available. DNA sequencers can only read a maximum number of sequences. As a result, the reads obtained by the sequencer constitute a random sample of the total number of DNA sequences and, thus, are compositional (Gloor et al. 2017). ASV counts were transformed using a clr transformation (Aitchison 1982) to obtain scale-invariant values, avoiding the compositional effect. 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 et al. 2016). A permutational multivariate analysis of variance was performed to assess the effect of the experimental design on the communities, 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 generalized mixed model using the lme4 R package v1.1-27.1 (Bates et al. 2015). *Z. tritici* qPCR analysis was also fitted in a generalized mixed model using lme4.

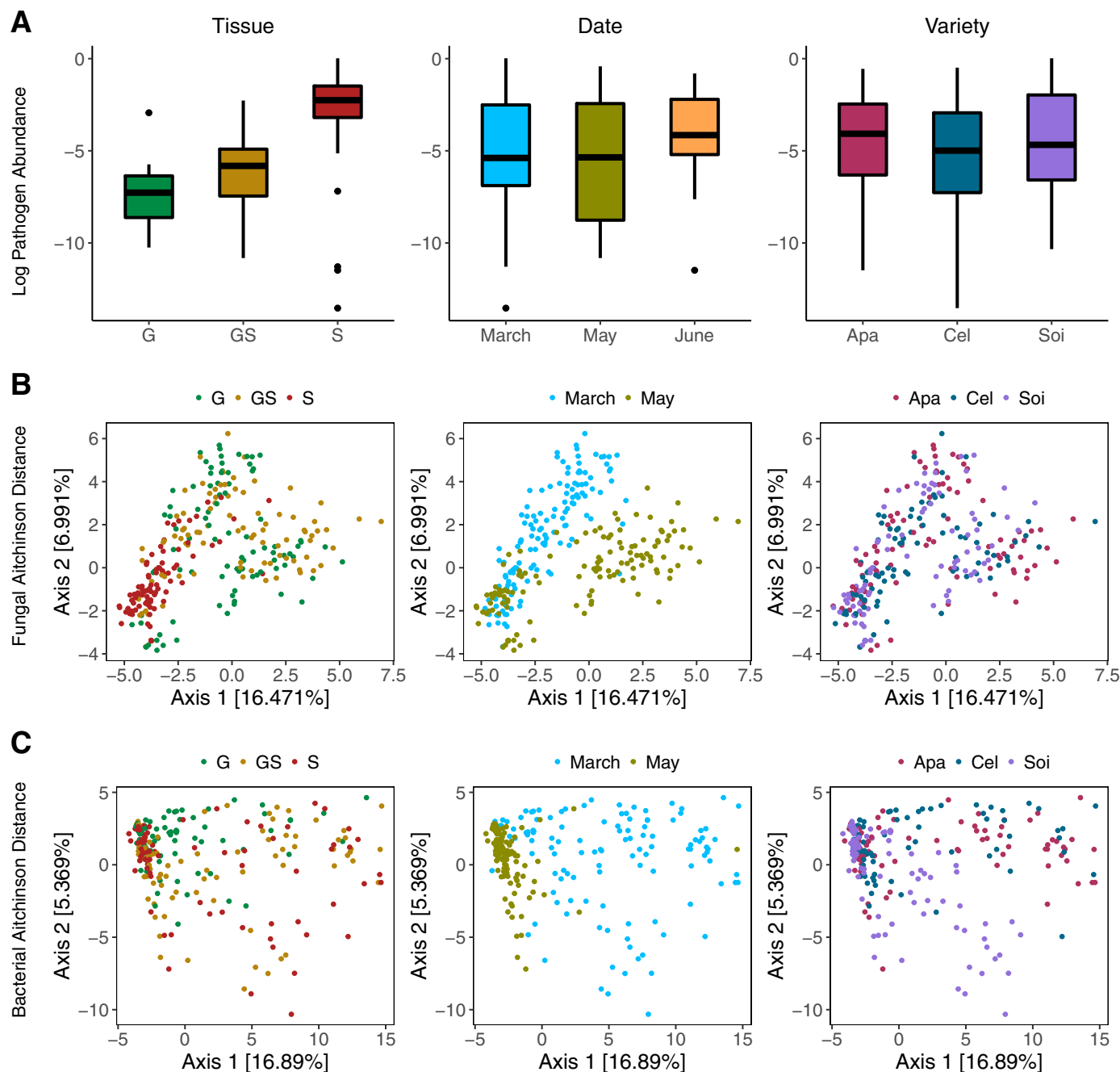
## RESULTS

This article provides two sets of raw sequence 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 Data Availability). 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 assignment, and a metadata table. The raw ASV tables also include the positive and negative 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, the visual assessment of symptoms, and the *Z. tritici* DNA concentration obtained by qPCR. The impacts of the visual

assessment of symptoms, the date, and the variety on the *Z. tritici* DNA concentration were significant (Fig. 1A; Supplementary Table S1). Although the varieties APA and SOI are more susceptible to STB than CEL, the difference in the concentration of *Z. tritici* DNA between them were limited, which is not surprising because sampling was based on similar leaf symptom criteria (G, GS, and S). The tables showing the change in number of reads in each sample during the bioinformatic process are also provided in the Dataverse files.

**Fungal communities.** All leaf samples ( $n = 360$ ) 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 ASV inference process identified an average of 28,178 high-quality sequences per sample distributed in 2,821 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. Three samples did not generate enough sequences to assign ASVs. The minimum number of reads in a sample was 20 and the maximum was 223,756. Several samples of weeds ( $n = 101$ ) growing close to the field were also sequenced using ITS primers. The



**Fig. 1.** Community composition of wheat leaves. **A**, Abundance of the pathogen (*Zymoseptoria tritici*) on wheat leaves measured using quantitative PCR. Principal coordinates analysis plots showing the similarity of **B**, fungal and **C**, bacterial communities from different samples. Ordination was performed using the Aitchison distance (Aitchison 1982). Plots were colored by the *Septoria tritici* blotch symptoms: leaf samples collected on asymptomatic leaves (G), green parts of a symptomatic leaf (GS), and symptomatic parts of a leaf (S); the sampling seasons of March, May, and June; and the wheat varieties sampled: Apache (Apa), Cellule (Cel), and Soissons (Soi).

bioinformatic process yielded an average 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 1,331 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.

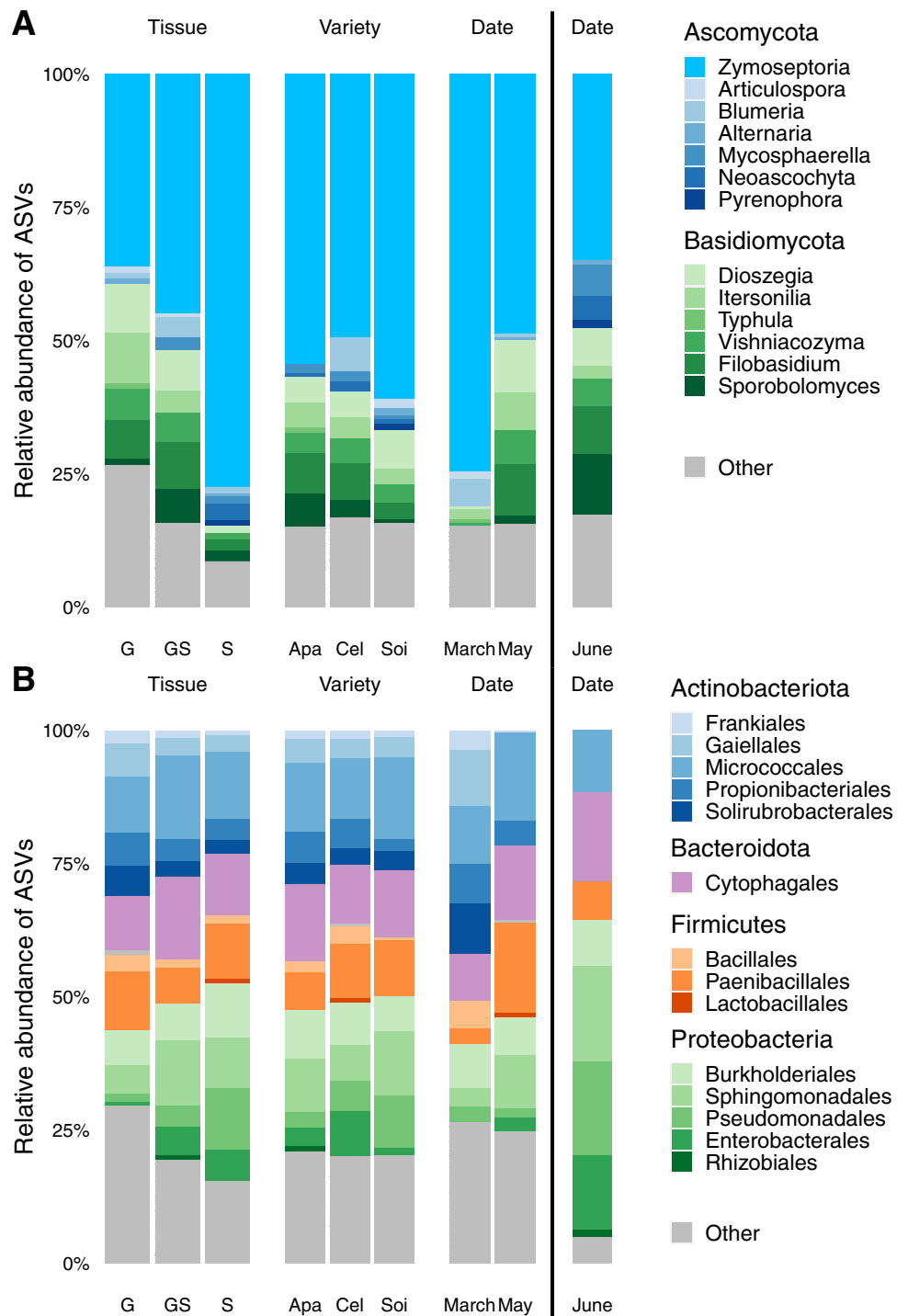
**Taxonomic composition.** For the wheat dataset, sequences assigned to Ascomycota represented 74% of the total counts, while sequences assigned to Basidiomycota represented 25% (Fig. 2A). As expected, ASVs assigned to the genus *Zymoseptoria* were the most abundant (60% of the sequences). *Zymoseptoria*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

the CEL cultivar, which is less susceptible because it carries the *Stb16q* resistance gene.

**$\alpha$ -Diversity.** Fungal community richness (number of species), diversity (Shannon diversity index), and evenness (inverse Simpson index) (Supplementary Fig. S1) differed significantly between dates and tissue health conditions (Supplementary Table S2). Wheat variety had a minor effect in all  $\alpha$ -diversity measures, being significant only for richness and diversity.

**$\beta$ -Diversity.** The composition of wheat foliar fungal communities (Fig. 1B) differed significantly among dates, varieties, and tissue health conditions (Table 1). Tissue was the most important factor, explaining 10% of the variance in the permutational analysis of

**Fig. 2.** Relative abundance of different taxa along different dates, wheat varieties, and leaf tissue conditions. **A**, Relative abundance of different fungal genera and **B**, relative abundance of different bacterial orders. ASV = amplicon sequence variant. Bar plots are separated by Septoria leaf blotch symptoms: leaf samples collected on asymptomatic leaves (G), green parts of a symptomatic leaf (GS), and symptomatic parts of a leaf (S); the wheat varieties sampled: Apache (Apa), Celule (Cel), and Soissons (Soi); and the sampling seasons of March, May, and June (June is separated by a bar because no G samples were available).



variance. Samples collected in June were not included in the analysis to avoid a potential bias caused by the absence of healthy leaves (G samples) at that time.

**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 reads and a maximum of 92,520 reads per sample. The ASV inference process identified a mean of 31,969 high-quality sequences shared between 12,349 unique ASVs in 350 samples. Ten samples did not generate enough sequences to assign ASV inference. The filtering process deleted contaminants and low abundant ASVs. The ASV table obtained had an average of 13,964 sequences per sample distributed between 1,495 ASVs and 340 samples. The minimum number of reads in a sample was 2 and the maximum was 71,051. The samples of weeds ( $n = 102$ ) 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 ASVs), followed by *Actinobacteria* (35%), *Bacteroidetes* (12%), and *Firmicutes* (11%) (Fig. 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 (Supplementary Fig. S1) differed significantly between tissue health conditions (Supplementary Table S3). Sampling date affected richness and evenness, while wheat variety only affected diversity.

**$\beta$ -Diversity.** The composition of the bacterial communities of wheat leaves (Fig. 1C) differed significantly among dates, varieties, and tissue conditions (Table 1). As with the fungal component of the community, date was the most important structuring factor, explaining some 8% of the variance in the permutational analysis of variance. Samples obtained in June were not included in the analysis to avoid a potential for bias caused by the absence of healthy leaf samples on that sample date.

**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. Although the three factors tested affected the community structure, the date of sampling exhibited the strongest effect. As expected, we found a relationship 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 tissue). Moreover, the large community overlap between samples from the three wheat varieties

**TABLE 1**  
Permutational multivariate analysis of variance assessing the effect of the experimental design on the communities<sup>a</sup>

Community	df	Sum of squares	$R^2$	$F$	$P$ value
<b>Fungi</b>					
Tissue	2	1,369.149	0.051	12.319	***
Date	2	2,856.615	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	3,510.502	0.131	1.316	***
Residual	285	15,838.138	0.591	–	–
Total	356	26,820.673	1	–	–
<b>Bacteria</b>					
Tissue	2	822.805	0.017	3.699	***
Date	2	6,153.514	0.126	27.664	***
Variety	2	1,189.58	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	1,918.814	0.039	4.313	***
Tissue × Date × Variety	6	909.533	0.019	1.363	**
Tissue × Date × Variety × Plot	48	6,775.259	0.138	1.269	***
Residual	268	29,806.437	0.608	–	–
Total	339	49,007.741	1	–	–

<sup>a</sup> 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 amplicon sequence variant tables separately using the R vegan package (Oksanen et al. 2019). Factors were Tissue, corresponding to the presence of symptoms caused by *Zymoseptoria tritici*; Date, corresponding to the sampling dates; and Variety, corresponding to the wheat varieties sampled. Plots were considered nested to the wheat variety. Asterisks \*, \*\*, and \*\*\* indicate  $P$  values < 0.05, 0.01, and 0.001, respectively.

(Fig. 1B and C) suggests that microbiomes are similar despite their variation in susceptibility to STB. These findings have implications for the concept that variation in the microbiome could be relevant to define and optimize biological control of STB, taking into account the impact of the wheat varieties. The preliminary analysis of this wheat dataset confirms the effectiveness of the sampling strategy and the need for further in-depth investigations. For instance, cooccurrence 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 noncrop plants within and in the margins of the field.

**Data availability.** The sequence datasets were deposited in the NCBI Sequence Read Archive in Bioproject PRJNA803042 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA803042/>). 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.

## ACKNOWLEDGMENTS

We thank all members of the Consortium Biocontrôle for their support for the project; Céline Lalanne, Adline Delcamp, Christophe Boury, and all other members of PGTB for their support in molecular biology; Gaétan Burgaud and Frédéric Garabetian for kindly providing marine strains used as positive controls; and Frédéric Barraquand, Stéphane Robin, Charlie Pauvert, and Andreas Makiola for helpful discussions at the beginning of the project.

## LITERATURE CITED

- Abarenkov, K., Zirk, A., Piirmann, T., Pöhönen, R., Ivanov, F., Nilsson, R. H., and Kõljalg, U. 2021. UNITE general FASTA release for eukaryotes 2. UNITE Community. <https://unite.ut.ee/repository.php>
- Aitchison, J. 1982. The statistical analysis of compositional data. *J. R. Stat. Soc. B* 44:139-160.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:1-48.
- Bengtsson-Palme, J., Hartmann, M., Eriksson, K. M., Pal, C., Thorell, K., Larsson, D. G. J., and Nilsson, R. H. 2015. METAXA2: Improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol. Ecol. Resour.* 15:1403-1414.
- Błaszczyk, L., Salamon, S., and Mikołajczak, K. 2021. Fungi inhabiting the wheat endosphere. *Pathogens* 10:1288.
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13:581-583.
- Cao, Q., Sun, X., Rajesh, K., Chalasani, N., Gelow, K., Katz, B., Shah, V. H., Sanyal, A. J., and Smirnova, E. 2021. Effects of rare microbiome taxa filtering on statistical analysis. *Front. Microbiol.* 11:607325.
- Chaudhry, V., Runge, P., Sengupta, P., Doehlemann, G., Parker, J. E., and Ke-men, E. 2020. Shaping the leaf microbiota: Plant-microbe-microbe interactions. *J. Exp. Bot.* 72:36-56.
- Chelius, M. K., and Triplett, E. W. 2001. The diversity of archaea and bacteria in association with the roots of *Zea mays*. *Microb. Ecol.* 41:252-263.
- Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., and Callahan, B. J. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6:1-14.
- Duvivier, M., Dedeurwaerder, G., De Proft, M., Moreau, J.-M., and Legrève, A. 2013. Real-time PCR quantification and spatio-temporal distribution of airborne inoculum of *Mycosphaerella graminicola* in Belgium. *Eur. J. Plant Pathol.* 137:325-341.
- Fones, H., and Gurr, S. 2015. The impact of Septoria tritici blotch disease on wheat: An EU perspective. *Fungal Genet. Biol.* 79:3-7.
- Frøsløv, T. G., Kjølner, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., and Hansen, A. J. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat. Commun.* 8:1188.
- Galan, M., Razzauti, M., Bard, E., Bernard, M., Brouat, C., Charbonnel, N., Dehne-Garcia, A., Loiseau, A., Tatar, C., Tamisier, L., Vayssier-Taussat, M., Vignes, H., and Cosson, J.-F. 2016. 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in Wildlife. *mSystems* 1:e00032-16.
- Gardes, M., and Bruns, T. D. 1993. ITS primers with enhanced specificity for basidiomycetes—Application to the identification of mycorrhizae and rusts. *Mol. Ecol.* 2:113-118.
- Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. 2017. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* 8:2224.
- Kerdraon, L., Barret, M., Laval, V., and Suffert, F. 2019. Differential dynamics of microbial community networks help identify microorganisms interacting with residue-borne pathogens: The case of *Zymoseptoria tritici* in wheat. *Microbiome* 7:125.
- Laforest-Lapointe, I., Paquette, A., Messier, C., and Kembel, S. W. 2017. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* 546:145-147.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17:10-12.
- McMurdie, P. J., and Holmes, S. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217.
- Morais, D., Gélisse, S., Laval, V., Sache, I., and Suffert, F. 2016. Inferring the origin of primary inoculum of *Zymoseptoria tritici* from differential adaptation of resident and immigrant populations to wheat cultivars. *Eur. J. Plant Pathol.* 145:393-404.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. 2019. vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>
- Orellana-Torres, C., Vidal, T., Boixel, A.-L., Gélisse, S., Saint-Jean, S., and Suffert, F. 2022. 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.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F. O. 2012. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41:D590-D596.
- R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Redford, A. J., Bowers, R. M., Knight, R., Linhart, Y., and Fierer, N. 2010. The ecology of the phyllosphere: Geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environ. Microbiol.* 12:2885-2893.
- Röttgers, L., and Faust, K. 2018. From hairballs to hypotheses biological insights from microbial networks. *FEMS Microbiol. Rev.* 42:761-780.
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., and Consortium, F. B. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc. Natl. Acad. Sci. U.S.A.* 109:6241-6246.
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73:5261-5267.
- White, T. J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-322 in: *PCR Protocols: A Guide to Methods and Applications*. M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. H. White, eds. Academic Press, San Diego, CA.
- Wickham, H., Navarro, D., and Pedersen, T. L. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.