FEMS Microbiology Ecology, 2025, 101, fiaf004
DOI: 10.1093/femsec/fiaf004
Advance access publication date: 8 January 2025
Research Article

Spinach seed microbiome characteristics linked to suppressiveness against *Globisporangium ultimum* damping-off

Makrina Diakaki[1,2,*], Beatriz Andreo Jimenez[1], Ezra de Lange[1], Patrick Butterbach[3], Liesbeth van der Heijden[4], Jürgen Köhl[1], Wietse de Boer[2,5], Joeke Postma[1]

[1]Wageningen Plant Research, Wageningen University and Research, 6708 PB Wageningen, The Netherlands
[2]Soil Biology Group, Wageningen University and Research, 6708 PB Wageningen, The Netherlands
[3]Germains Seed Technology, 1601 MH Enkhuizen, The Netherlands
[4]Bejo Zaden B.V., 1749 CZ Warmenhuizen, The Netherlands
[5]Department of Microbial Ecology, Netherlands Institute of Ecology, 6708 PB Wageningen, The Netherlands
*Corresponding author. Droevendaalsesteeg 1, 6700 AA Wageningen, The Netherlands. E-mail: makrina.diakaki@wur.nl
Editor: Angela Sessitsch

Abstract

Recently we demonstrated that the seed microbiome of certain spinach (*Spinacia oleracea*) seed lots can confer disease suppression against *Globisporangium ultimum* damping-off (previously known as *Pythium ultimum*). We hypothesized that differences in the microbial community composition of spinach seed lots correlate with the levels of damping-off suppressiveness of each seed lot. Here, we show that a large proportion of variance in seed-associated bacterial (16S) and fungal (Internal Transcribed Spacer 1) amplicon sequences was explained by seed lot identity, while 9.8% of bacterial and 7.1% of fungal community variance correlated with disease suppression. More specifically, a higher relative abundance of basidiomycetous dimorphic yeasts such as *Vishniacozyma*, *Filobasidium*, and *Papiliotrema* and of the bacterial genus *Massilia* was a key feature of suppressive seed microbiomes. We suggest that the abundance of these genera is indicative of seed lot suppressive potential. Seed processing and treatment can become more targeted with indicator taxa being used to evaluate the presence of beneficial seed-associated microbial functions. This process, in turn, could contribute to the sustainable management of seedling diseases. Finally, this study highlights the ubiquity of yeasts in spinach seed microbiota and their potential beneficial roles for seed health.

Keywords: seed microbiome; bacteria; fungi; yeasts; *Pythium ultimum*; *Spinacia oleracea*

Introduction

Recent research has highlighted the role of seed microbiota on plant health and crop yield (Simonin et al. 2022, Verma and White 2019). With plant pests and diseases leading to substantial yield losses yearly (Mesterházy et al. 2020), understanding how seed microbiota can mitigate such losses offers a promising new asset for the production of healthy seeds and, consequently, for global food security (Husenov et al. 2021, Mcguire and Sperling 2011). This is of particular interest especially as several chemical seed-applied plant protection products are expected to be phased out in the EU.

A large-scale screening of 260 seed lots from seven different crops recently demonstrated that the seed microbiome of certain spinach (*Spinacia oleracea*) and beetroot (*Beta vulgaris*) seed lots can harbour beneficial microorganisms able to suppress seedling-stage disease (Diakaki et al. 2022). The pathogen against which suppressiveness was detected was *Globisporangium ultimum*, previously known as *Pythium ultimum*. This is an oomycete able to infect crops of high importance such as maize, soybean and wheat as well numerous vegetable crops, including members of the *Solanaceae*, *Brassicaceae*, and *Cucurbitaceae* families (Rai et al. 2020). With *G. ultimum* able to cause damping-off, plants are confronted with this pathogen during a vulnerable stage in plant growth, namely, the onset of germination, when seeds are still mainly colonized by seed microbiota. The process of germination is initiated when water is taken up by the seed (imbibition), after which the embryo physiologically prepares for emergence, before finally the embryonic axis (typically the radicle) protrudes from the structures surrounding it (Bewley et al. 2013). While germination represents an opportunity for infection (Rai et al. 2020), it is also a turning point for the assembly of plant microbiota (Barret et al. 2015, Chesneau et al. 2020). Evidence suggests that despite the importance of soil-derived microbiota (Rochefort et al. 2021, Wolfgang et al. 2020), seed-associated bacteria and fungi are also transmitted to young plants (Johnston-Monje et al. 2021, Nelson 2018) and can be essential for plant health (Jack and Nelson 2018, Matsumoto et al. 2021). Our previous research gave further support to the evidence of the importance of seed-associated microbiota for seed health, especially during germination (Diakaki et al. 2022).

To better understand the characteristics of seed microbiota that are suppressive to *G. ultimum*, we focused on eight spinach seed lots differing in damping-off suppressiveness and performed a second spinach—*G. ultimum* bioassay, confirming our previous findings (Diakaki et al. 2022). We amplified and sequenced microbial DNA extracted from seeds and analysed the relative abun-

Received 18 October 2024; **revised** 23 December 2024; **accepted** 7 January 2025

dance and taxonomical composition of the bacterial and fungal fractions of seed microbiota based on 16S rRNA gene and Internal Transcribed Spacer 1 (ITS1) amplicons, respectively. We hypothesized that the microbial communities of the eight seed lots included in this study differ in taxonomical composition and that these differences correlate with the levels of damping-off suppressiveness of each seed lot. Consequently, we addressed the following research questions: (i) How do the seed microbiota of different spinach seed lots differ in taxonomical composition? (ii) Do these differences correlate with seed microbiome suppressive potential against *G. ultimum*? And, lastly, (iii) are there specific taxa which are differentially abundant in highly suppressive seed lots? Determining the key taxa of a suppressive seed microbiome could assist the seed industry to develop indicators for suppressiveness and safeguard them during seed production and processing.

Materials and methods

Seed lots

This study focused on eight spinach seed lots coded A, B, C, D, E, F, G, and H, with each seed lot representing a unique genotype (cultivar). Seed lots were provided by Nunhems Netherlands B.V. (BASF), Pop Vriend Seeds B.V., Bejo Zaden B.V., and Sakata Vegetables Europe S.A.S. They were selected by these companies directly after harvest and were stored under uniform conditions (12°C, 40% RH). Seed lots A, B, C, D, E, and F were produced in Location 1, while seed lot G in Location 2 and seed lot H in Location 3, with each location representing a different country and fields of production being at least 3 km distant from one another for all seed lots. The seed lots have been selected based on the results of the spinach—*G. ultimum* bioassay reported by Diakaki et al. (2022) in which 40 spinach seed lots were tested for microbiome-associated suppressiveness. The eight seed lots represent a diverse range of microbiome-associated suppressiveness and have been coded consecutively, with seed lot A being the most suppressive and seed lot H the least. We created the grouping factor 'performance' to describe the suppressiveness of these seed lots. We considered seed lots A, B, C, and D to have a high level of suppressive potential (high performance) and seed lots E, F, G and H a low level (low performance).

Spinach—*Globisporangium ultimum* bioassays

A year after completing the spinach—*G. ultimum* bioassay reported by Diakaki et al. (2022), a second bioassay was performed to confirm the reproducibility and persistence of our findings for a subset of 16 seed lots. The set-up of the two bioassays was identical, with the exception of the first including four replicates of 62 seeds each, while the second had three replicates of 48 seeds per treatment. Detailed information can be found in the supplementary materials and methods of Diakaki et al. (2022).

Briefly, a part of the spinach seeds was treated with steam vacuum pasteurization by Bejo Zaden B.V. as a proxy for partially eliminating the seed microbiota. Seeds were treated a few weeks before the onset of each bioassay. Subsequently, treated and non-treated spinach seeds were challenged with *G. ultimum*. The bioassays used field top soil that was collected in the area of Breda, The Netherlands, from a field that was naturally infected with *G. ultimum*. Every replicate comprised an individual container. The seeds were sown on the naturally infected soil and covered by a 1-cm layer of potting soil. The bioassays were done in the same growing chamber (50% humidity, 20°C day/15°C night, 12 h light/12 h darkness) and treatments were randomized. The number of germinated seedlings as well as the number of seedlings presenting post-emergence infection symptoms were scored 6 and 7 days after sowing.

Germination and emergence tests confirmed that no physiological side-effects were induced in the seeds of the selected seed lots due to steam vacuum pasteurization. The germination test conducted along with the first bioassay followed the protocols established by the International Seed Testing Association (ISTA 2020) as described in the supplementary materials and methods of Diakaki et al. (2022). The emergence test conducted along with the second bioassay was done by sowing seeds (three replicates of 48 seeds per seed lot) in the same naturally infected soil after a 3-h steaming pasteurization period at 70°C, to eliminate *G. ultimum*, followed by a 'recovering period' of 3 weeks at 15°C in plastic breathable bags, in the dark. The number of emerged seedlings was then recorded 6 and 7 days after sowing.

DNA metabarcoding

Seed samples used for metabarcoding were placed and subsequently stored in the freezer (−20°C) at the time when the first bioassay was taking place. For every seed lot, DNA extraction and PCR was conducted for ten samples of ∼70–80 mg of non-treated seeds. The samples were pooled in pairs after PCR to make up the five biological replicates that were sequenced per seed lot. Samples were processed and sequenced in two sets: samples from seed lots A, C, G, and H were processed first (set 1), followed by seed lots B, D, E, and F (set 2).

Sample preprocessing prior to DNA extraction followed the protocol of Köhl et al. (2024). The Precellys Evolution (Bertin Technologies) was used to bead-beat ∼70–80 mg of seeds with a 6.35 mm steel bead (BioSpec) inside reinforced 2 ml screw-cap tubes at 5,000 r/m for 30 s, followed by a 20-s pause and bead-beating for 30 s a second time. Bead-beating for 30 s took place a third time when seeds remained partially intact to ensure producing homogenously powdered samples. Samples were then freeze-dried overnight, followed by transferring ∼15 mg of lyophilized powder per sample to a well of a 96-well block. The block was placed in liquid nitrogen for 30 s followed by adding 90 mg of 1.0 mm silicon carbide beads (BioSpec) to each well and bead-beating the samples using TissueLyser II (Qiagen) at 30 Hertz for 30 s, inverting the block and bead-beating a second time at 30 Hertz for 30 s. DNA extraction then took place using the PurePrep Seed Kit (Molgen), the Kingfisher Flex robot (Thermofisher) and following the respective manufacturer protocols. DNA extraction included five negative controls for set 1 and three for set 2. We had initially used the sbeadex™ maxi plant kit (LGC) for DNA extraction but needed to discard the samples generated with this. This kit is commercialized for conducting plant DNA extractions and after repeatedly finding remnants of bacterial DNA in its reagents, we discourage its use for microbiome studies.

The Q5 Hot Start High-Fidelity DNA polymerase (New England Biolabs) PCR kit was used for DNA amplifications. Two sets of primers were used, namely primers for the V4 region of the 16S rRNA gene: 515F (Parada) 5′- GTGYCAGCMGCCGCGGTAA -3′ and 806R (Apprill) 5′- GGACTACNVGGGTWTCTAAT -3′ and primers for ITS1: ITS1F 5′- CTTGGTCATTTAGAGGAAGTAA -3′ and ITS2 5′- GCTGCGTTCTTCATCGATGC -3′ (Apprill et al. 2015, Gardes and Bruns 1993, Parada et al. 2016, White et al. 1990). Amplifying the 16S rRNA gene and ITS1 aimed at assessing the bacterial and fungal fractions of the seed microbiota, respectively. Amplification of plant DNA in the 16S rRNA gene PCR was blocked with the use of PNA clamps, namely the anti-mitochondrial PNA (mPNA) 5′-GGCAAGTGTTCTTCGGA-3′ and the anti-plastid PNA (pPNA) 5′-GGCTCAACCCTGGACAG-3′ (PNA Bio) (Lundberg et al. 2013). DNA

extracts were diluted 1:5 and ∼20 ng DNA were used per 50 µl PCR assay for both the 16S rRNA gene and the ITS1 PCRs.

For the 16S rRNA gene PCR of sample set 1, every assay comprised 2.5 µl (1:5 diluted) DNA extract, 2.5 µl per primer (10 µM), 2 µl dNTPs (5 mM), 0.5 µl Q5 High-Fidelity DNA polymerase, 5 µl mPNA (50 µM), 5 µl pPNA (50 µM), 20 µl water, and 10 µl Q5 Reaction buffer (5×). The thermal cycling started with heating the samples at 98°C for 3 min, followed by 26 cycles of 95°C (30 s), 75°C (10 s), 50°C (30 s), and 72°C (30 s), before final elongation phase of 1 min at 72°C. For the ITS1 PCR of set 1, every assay comprised 2.5 µl (1 : 5 diluted) DNA extract, 2.5 µl per primer (10 µM), 2 µl dNTPs (5 mM), 0.5 µl Q5 High-Fidelity DNA polymerase, 30 µl water, and 10 µl Q5 Reaction buffer (5×). The thermal cycling started with heating the samples at 98°C for 3 min, followed by 28 cycles of 95°C (30 s), 75°C (10 s), 52°C (30 s), and 72°C (30 s), before a final elongation phase of 1 min at 72°C. The PCRs of sample set 2 only differed in the assay composition by including 4 µl (1 : 5 diluted) DNA extract and 18.5 µl water (16S rRNA gene) or 28.5 µl water (ITS1). Each PCR included a negative control sample amounting to four negative PCR control samples in total.

The PCRs of set 2 also included positive control samples in three replicates. For bacteria, the commercial ZymoBIOMICS™ Microbial Community Standard (D6300) (Zymo Research) was used. This commercial DNA standard includes equimolar amounts of DNA from *Listeria monocytogenes*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, *Escherichia coli*, *Salmonella enterica*, *Lactobacillus fermentum*, *Enterococcus faecalis*, and *Staphylococcus aureus*. For fungi, a DNA standard was prepared in-house by mixing purified full-length ITS DNA amplicons from six known pure fungal isolates. This DNA standard included DNA ∼ 2×10^8 amplicon copies/µl per isolate from *Monographella nivalis*, *Verticillium dahliae*, *Cladosporium cladosporioides*, *Vishniacozyma victoriae*, *Sporobolomyces roseus*, and *Agaricus bisporus*. Amplification of the ZymoBIOMICS™ Microbial Community Standard (D6300) replicates followed the 16S rRNA gene PCR settings, with a total input of 20 ng DNA per 50 µl PCR assay. Replicates of the in-house DNA standard mixture were amplified following the ITS1 PCR settings with the exception of comprising 21 cycles and using 1.5 ng DNA as input per 50 µl PCR assay. To verify that the preparation of sample sets 1 and 2 was comparable, one biological replicate per seed lot from set 1 was also amplified and sequenced along with the samples of set 2.

After PCR, amplicons from every seed lot were pooled in pairs to make up five biological replicates for sequencing. These pooled samples and the control samples were sent to IGA Technology Services Srl, with the exception of the four negative PCR control samples, all of which resulted in no amplification. The sequencing company purified the samples using 2× XP Beckman beads (Beckman Coulter), and sequenced them using Illumina MiSeq (2 × 300 PE). After purification, the negative controls of set 1 contained insufficient amounts of DNA and could not be sequenced. The negative controls of set 2 were pooled into a single sample of 150 µl per library (16S rRNA gene and ITS1) prior to purification and were sequenced. Information on the controls of the study can be found in Supplementary File S1.

Sequencing data processing (QIIME2)

Sequenced reads were processed using QIIME2 (version 2021.8) (Bolyen et al. 2019). Reads that were shorter than 100 bp were removed using the cutadapt plugin (Martin 2011) after which the remaining reads were demultiplexed, merged (16S reads only) and quality-filtered using the DADA2 plugin (Callahan et al. 2016). They were then classified as amplicon sequence variants (ASVs) and taxonomically annotated using the feature-classifier of QIIME2 based on Naïve Bayes algorithm (Bokulich et al. 2018). The annotation for the 16S rRNA gene amplicons was based on the Silva database (v.132) (Quast et al. 2012), while for the ITS1 amplicons, it was based on the UNITE database (version 8, May 2021) (Nilsson et al. 2019). Taxa classified as Archaea and Eukaryota were removed from the bacterial dataset. For both the bacterial and the fungal datasets, taxa being represented by 10 reads or less in the entire ASV table were removed, together with taxa that were not identified at the phylum level. The resulting bacterial and fungal ASV tables were then imported in R using the qiime2R (Bisanz 2018) and phyloseq (McMurdie and Holmes 2013) R packages.

Statistical analysis and data visualization

All further data handling and the statistical analysis of the datasets were performed with R version 4.2.1 (R Core Team 2024). Data were visualized with the ggplot2 R package (Wickham 2016). All data recorded in the bioassays and germination tests were count data (number of germinated, emerged or diseased seedlings) and were analysed using generalized linear models ($\alpha = 0.05$) with the glm function of the multcomp R package (Hothorn et al. 2008). The dispersion of count data was examined using the testDispersion function of the DHARMa R package (Hartig 2019). In order to account for data over- or under-dispersion, a quasibinomial distribution was used in the generalized linear models. False discovery rate (FDR) was controlled using the Benjamini–Hochberg procedure (BH) at level $\alpha = 0.05$, given that the analysis comprised multiple comparisons. While in Diakaki et al. (2022) the number of pre- and post-emergence damped-off seedlings were treated as two separate disease variables, here they were summed per replicate and analysed as a single variable.

The results of the first bioassay were used as metadata in the analysis of the sequencing data, since the seed samples used for metabarcoding were placed in the freezer (−20°C) at the time when that bioassay was taking place.

With regards to the sequencing data, the rarefy_even_depth function of the phyloseq R package (McMurdie and Holmes 2013) was used for normalizing the ASV tables by rarefaction at a sampling depth of 12 717 reads (bacterial dataset) and 23 979 reads (fungal dataset). This resulted in the bacterial dataset being reduced to 37 samples, after the removal of samples which failed to meet the selected sampling depth threshold (two samples of seed lot E and one of seed lot G). All 40 samples of the fungal dataset were conserved. Rarefied ASV tables were used for calculating alpha diversity metrics, namely species richness and Shannon diversity index. For all other types of analyses, read counts per ASV were transformed into proportions (relative abundance), as a normalization method to adjust for unequal library size and retain all ASVs, using the transform function of the microbiome R package (Lahti and Shetty 2018). One of the five replicates of seed lot E was excluded from all analyses due to its low library size (number of reads). The following steps of the statistical analysis were completed for both datasets separately. Where necessary, FDR was controlled using the Benjamini–Hochberg procedure (BH) at level $\alpha = 0.05$. Where possible, the factor location was constrained since it is known that terroir is a major component shaping seed microbiota (Klaedtke et al. 2016, Morales Moreira et al. 2021b) and as such, was expected to mask the factors of interest, namely seed lot, disease suppression and performance.

Species richness and Shannon diversity index were calculated per seed lot using the estimate_richness function of the phyloseq R package (McMurdie and Holmes 2013). The Kruskal–Wallis test of R (BH; $\alpha = 0.05$) was then used to test whether these two diver-

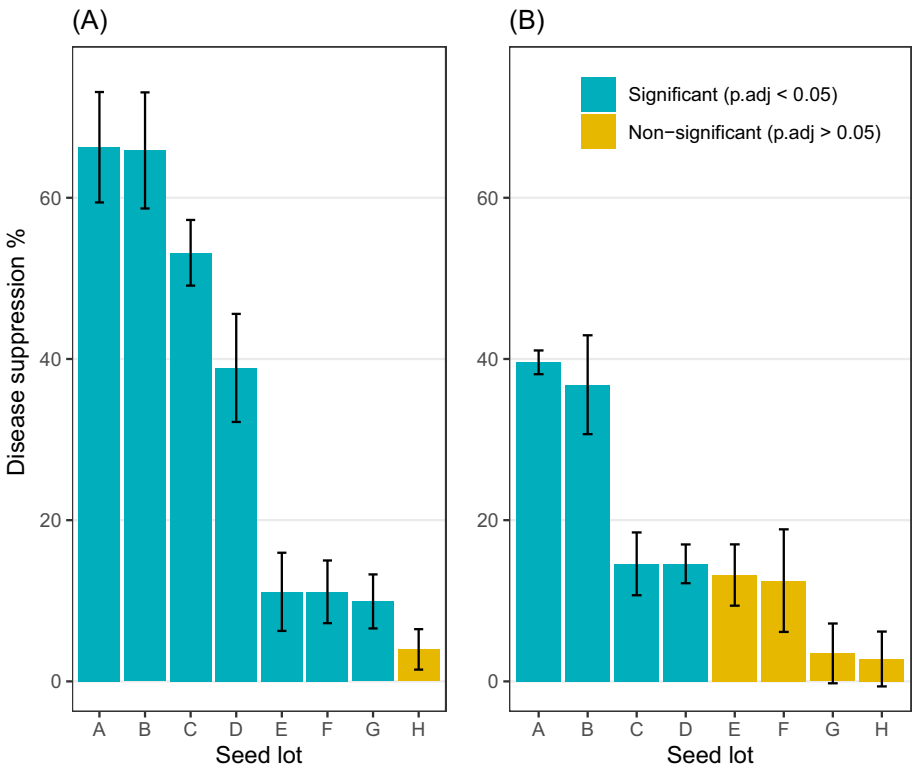

Figure 1. Disease suppression per seed lot in spinach—*Globisporangium ultimum* bioassays. The y-axis (disease suppression %) represents the mean difference in damping-off infection % between treated and non-treated seed samples per seed lot. Seed lots for which this difference is significant (generalized linear model, $\alpha = 0.05$) appear in blue. Data are shown for the first [A; based on data from Diakaki et al. (2022)] and second bioassay (B). False discovery rate (FDR) was controlled using the Benjamini–Hochberg procedure ($\alpha = 0.05$). For seed lot C in the second bioassay, 100% of treated seeds were diseased in all replicates; the colour of the respective bar has been manually adjusted to blue since the lack of variation interfered with the statistical analysis.

sity metrics differed per seed lot. The Pearson's correlation test (BH; $\alpha = 0.05$) of R was used to evaluate the correlation between each of the diversity metrics and disease suppression.

For categorizing genera as core or rare, the tax_glom function of phyloseq R package (McMurdie and Holmes 2013) was first used to agglomerate taxa at the genus level. The core_members and rare_members functions of the microbiome R package (Lahti and Shetty 2018) were then used to extract core and rare taxa, respectively. The same was done for ASVs without an agglomeration step. Similarly to Kim et al. (2023), we defined taxa as core when being present in at least 80% of the samples and rare when being present in no more than 20% of the samples. Prevalence was calculated per ASV as the percentage of samples in which the ASV is present.

Ordination analysis was performed based on the Bray–Curtis and the Jaccard dissimilarity matrices (non-metric multidimensional scaling; NMDS), using the ordinate function of the phyloseq R package (McMurdie and Holmes 2013). The Bray–Curtis distance metric accounts for presence/absence and relative abundance data, while the Jaccard distance metric only for presence/absence data. Permutational multivariate analysis of variance (PERMANOVA) was performed using the adonis2 function of the vegan R package (Oksanen 2012) to explore the percentage of variance explained by the factors seed lot and disease suppression; the factor location was constrained in the models.

Relating to the identification of differentially abundant taxa between high and low-performing seed lots, three different methods were used, namely, a log-linear (natural log) model (ancombc2 function of the ANCOMBC R package) (Lin et al. 2022, Lin and Peddada 2020), a Wilcoxon rank-sum test (compare_groups function of the metacoder R package) (Foster et al. 2017) and a Wald test (DESeq function of the DESeq2 R package) (Love et al. 2014) (BH; $\alpha = 0.05$). This was decided since it has been shown that the numbers and sets of differentially abundant taxa can vary significantly when using different tools to identify them (Nearing et al. 2022). We especially highlight the results of the ANCOM-BC2 model over the other two, since we consider it more robust and suitable based on literature (Lin and Peddada 2024, Nearing et al. 2022). Additionally, the ANCOM-BC2 model, allowed the inclusion of the factor location, while the other methods could only incorporate one factor, namely performance. Lastly, the R package indicspecies was used to perform an indicator species analysis for the two performance groups (De Cáceres and Legendre 2009).

All datasets, including the raw sequencing data and scripts used for their analyses can be found in the 4TU database with doi 10.4121/03e9e2af-904e-4949-ac10-c5721fd6e0f1, which can be accessed at https://www.data.4tu.nl. The raw sequencing data used are also publicly available via the European Nucleotide Archive under project accession PRJEB77602.

Results

Spinach—*Globisporangium ultimum* bioassays

The eight seed lots selected for this study are characterized by different levels of suppressive potential, a trait that was evident and consistent in the results of both the first (Diakaki et al. 2022) and second bioassay. *Globisporangium ultimum* inoculum density and infectivity were not the same in the two bioassays since naturally infested soil was collected a year apart to be used as substrate. Seeds had also aged by one year. Fig. 1 presents the mean differ-

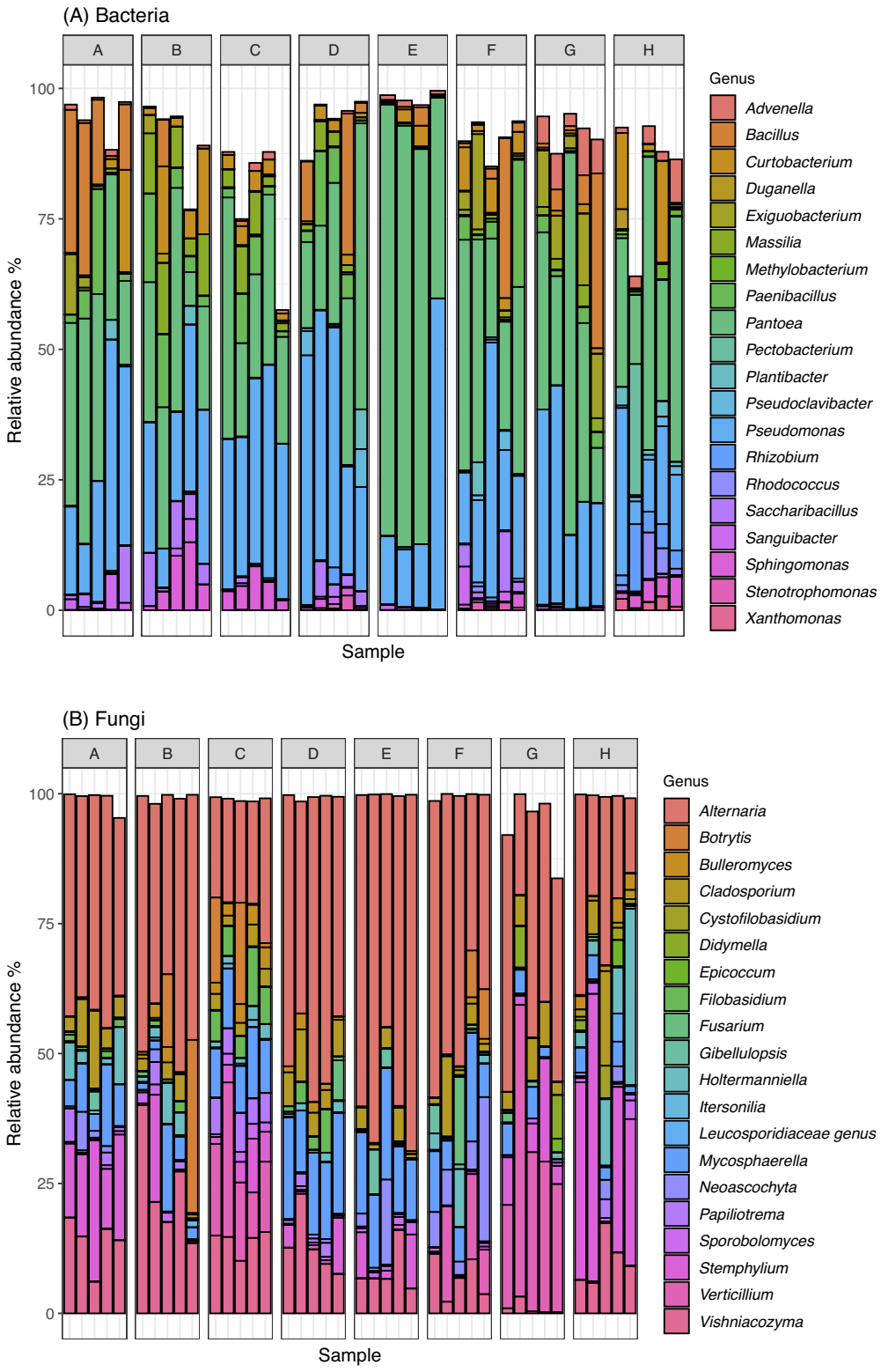

Figure 2. Top 20 bacterial (A) and fungal (B) genera across samples. The barplots show the relative abundance (%) of the 20 most abundant bacterial (A) and fungal (B) genera across samples.

ence in damping-off infection % between treated and non-treated seed samples per seed lot. Figure S1 (Supplementary File S2) presents the disease incidence in treated and non-treated seed samples separately. The larger the difference, the higher the suppressive potential of the seed microbiome of a seed lot. These differences were significant (generalized linear model, $\alpha = 0.05$) with the exception of seed lot H in both bioassays and seed lots E, F, and G in the second bioassay. This metric will be referred to as 'disease suppression' and is based on the values derived from the first bioassay. With regards to the grouping factor 'performance',

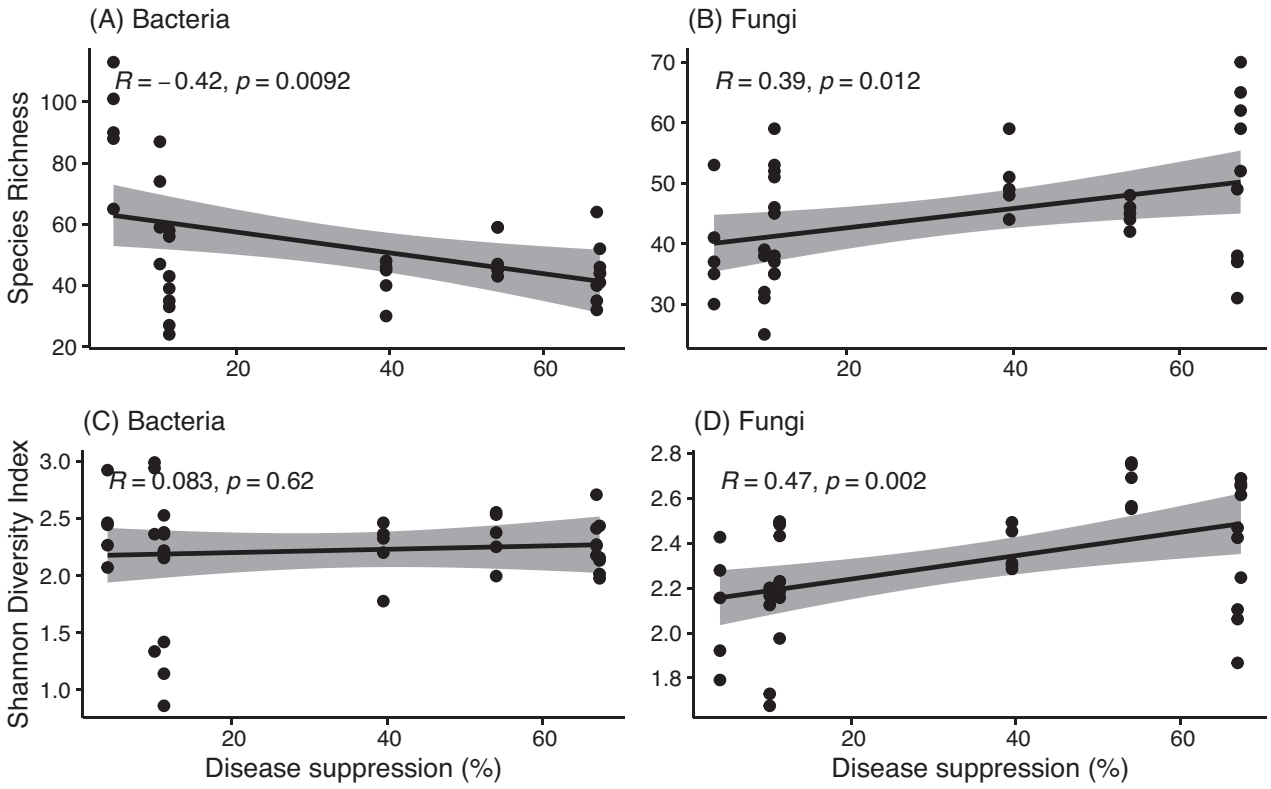

Figure 3. Pearson's correlation between alpha diversity metrics and seed lot disease suppression (%). (A) species richness for bacterial communities, (B) species richness for fungal communities, (C) Shannon diversity index for bacterial communities and (D) Shannon diversity index for fungal communities.

the disease suppression characterising high-performing seed lots ranged from 39% to 66%, while for low-performing seed lots, it ranged from 4% to 11%.

For seed lots A and B disease suppression was the highest in both the first (>65%) (Diakaki et al. 2022) and second (>36%) bioassay, while it was the lowest for seed lots E, F, G, and H (<14% in both bioassays). Disease suppression for seed lots C and D was 53% and 39%, respectively, in the first bioassay (Diakaki et al. 2022) and 15% in the second bioassay. It should be noted that neither location of cultivation nor company of production correlated with disease suppression (Kruskal–Wallis rank sum test, $\alpha = 0.05$); this was tested based on the results from all 40 seed lots in the first and all 16 seed lots in the second bioassay.

Sequencing data

Taxonomic profiling of the spinach seed microbiota

The bacterial dataset was described by an average of 27 770 high-quality reads per sample and a total of 343 ASVs, while the fungal dataset contained an average of 62 261 high-quality reads per sample and 273 ASVs.

For bacteria, most ASVs belonged to the class *Gammaproteobacteria* (70.39%), with smaller proportions belonging to *Bacilli* (14.37%) and *Actinobacteria* (10.98%) (Supplementary File S2: Figs. S2 and S3; Supplementary File S3). At the genus level, *Pantoea* (37.42%) and *Pseudomonas* (27.96%) were the most abundant genera (Fig. 2, Supplementary File S2: Fig. S3; Supplementary File S3). For fungi, most ASVs belonged to the class *Dothideomycetes* (67.10%), followed by *Tremellomycetes* (19.73%) and *Sordariomycetes* (9.17%) (Supplementary File S2: Figs. S2 and S3; Supplementary File S4), while *Alternaria* (40.88%) and *Vishniacozyma* (11.50%) were the most abundant genera (Fig. 2, Supplementary File S2: Fig. S3; Supplementary File S4).

For computing alpha diversity metrics, the number of reads was rarefied to 12 717 and 23 979 for the bacterial and fungal datasets, respectively. The number of ASVs retained was 333 bacterial and 273 fungal ASVs.

Alpha diversity differs per seed lot and correlates with disease suppression

For bacterial communities, species richness ranged from 24 to 113 ASVs per seed lot and differed significantly between lots (Kruskal-Wallis rank sum test, FDR-adjusted $P = 0.001$), while Shannon diversity index did not (Kruskal–Wallis rank sum test, FDR-adjusted $P = 0.106$) (Supplementary File S2: Fig. S4). Additionally, there was a moderate negative correlation between species richness and disease suppression (Pearson's correlation, $R = -0.42$, $P = 0.009$; Fig. 3). For fungal communities, species richness ranged from 25 to 70 ASVs per seed lot and differed significantly between lots (Kruskal–Wallis rank sum test, FDR-adjusted $P = 0.002$) (Supplementary File S2: Fig. S4). Shannon diversity index also differed significantly between seed lots (Kruskal–Wallis rank sum test, FDR-adjusted $P = 0.0003$) and ranged from 1.68 to 2.76 (Supplementary File S2: Fig. S4). Both species richness ($R = 0.39$, $P = 0.012$) and Shannon diversity index ($R = 0.47$, $P = 0.002$) correlated positively with disease suppression (Fig. 3). These Pearson's correlations were also moderate.

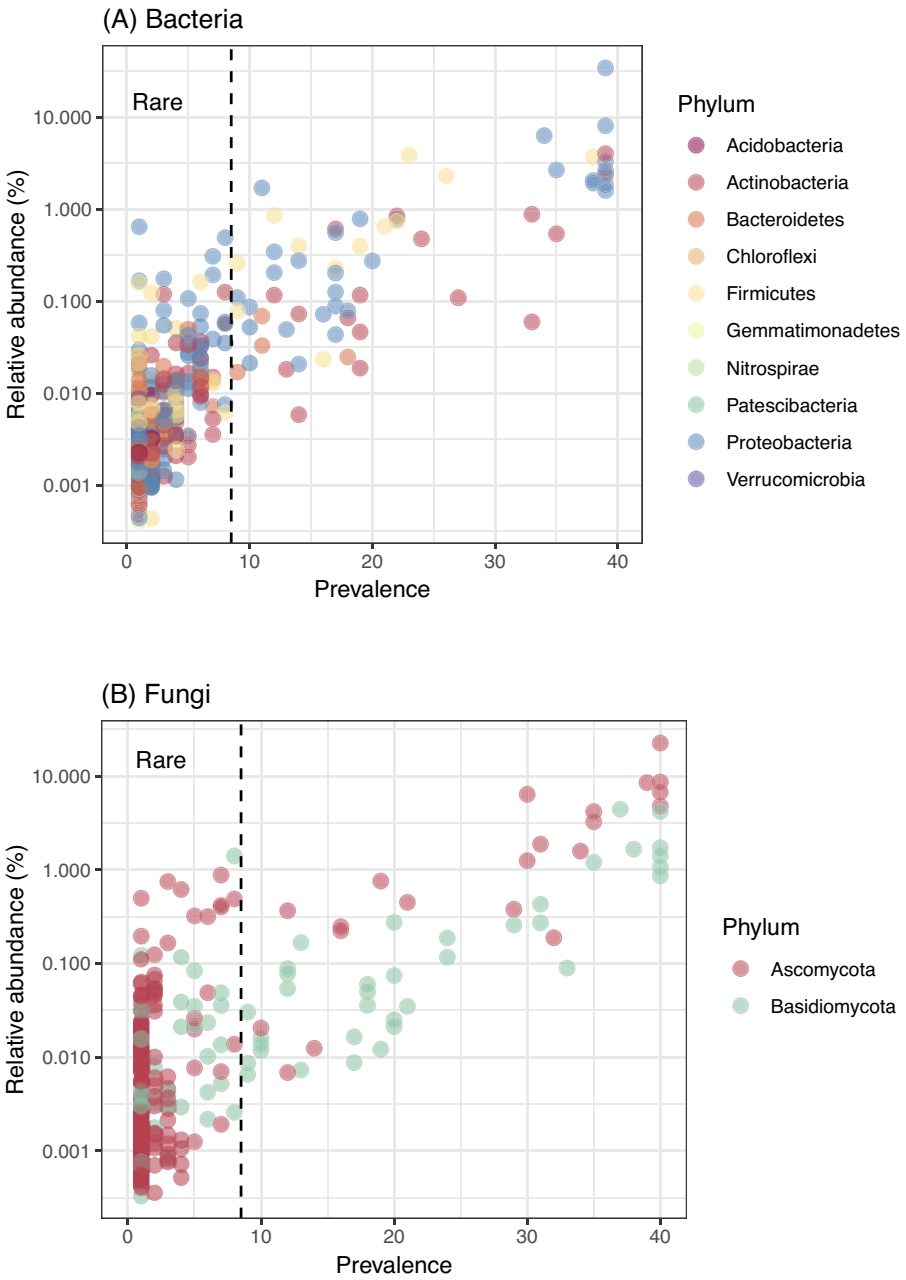

Figure 4. Relative abundance (%) and prevalence of bacterial (A) and fungal (B) ASVs. Prevalence represents the number of samples where a taxon is found. Each point represents an ASV. There are 39 samples in the bacterial dataset and 40 samples in the fungal dataset. Rare taxa (prevalence <20%) are found on the left of the black dashed line. When multiple points overlap, they appear darker. Point colour designates the phylum an ASV belongs to. Note: the y-axis is presented on a $\log_{10}$ scale.

Most taxa are rare while the seed lots also share core microbiota

The majority of the bacterial and fungal ASVs could be considered as rare based on their low prevalence (number of samples in which the ASV is present) and low relative abundance. From the total of 343 bacterial ASVs, 81% (280 ASVs) were rare (prevalence < 20%), with 32% (111 ASVs) only present in a single sample, while only 5% (16 ASVs) were found in >80% of the samples. At the same time, 85% of ASVs (290 ASVs) had a relative abundance of <0.1%, while 5% (16 ASVs) had a relative abundance >1% (Fig. 4). Similarly, from the 273 fungal ASVs, 79% (216 ASVs) were rare (prevalence < 20%), with 52% (141 ASVs) only present in a single sample, while only 7% (18 ASVs) were found in >80% of the samples. In addition, 82% of ASVs (224 ASVs) had a relative abundance of <0.1%, while 7% (19 ASVs) had a relative abundance >1% (Fig. 4). Lastly, we calculated that 56% of bacterial and 58% of fungal genera are only present in one seed lot.

A set of spinach seed core microbiota was identified. These were taxa that were present in 80% of the samples and were also present in at least one sample per seed lot. At the genus level, there were 11 core bacterial and 13 core fungal genera. The core bacterial genera in order of highest relative abundance were: *Pantoea*, *Pseudomonas*, *Paenibacillus*, *Curtobacterium*, *Massilia*, *Sphingomonas*, *Advenella*, *Sanguibacter*, *Rhizobium*, *Methylobacterium*, and *Rubrobacter* (Supplementary File S2: Fig. S5). The core fungal genera in order of highest relative abundance were: *Alternaria*,

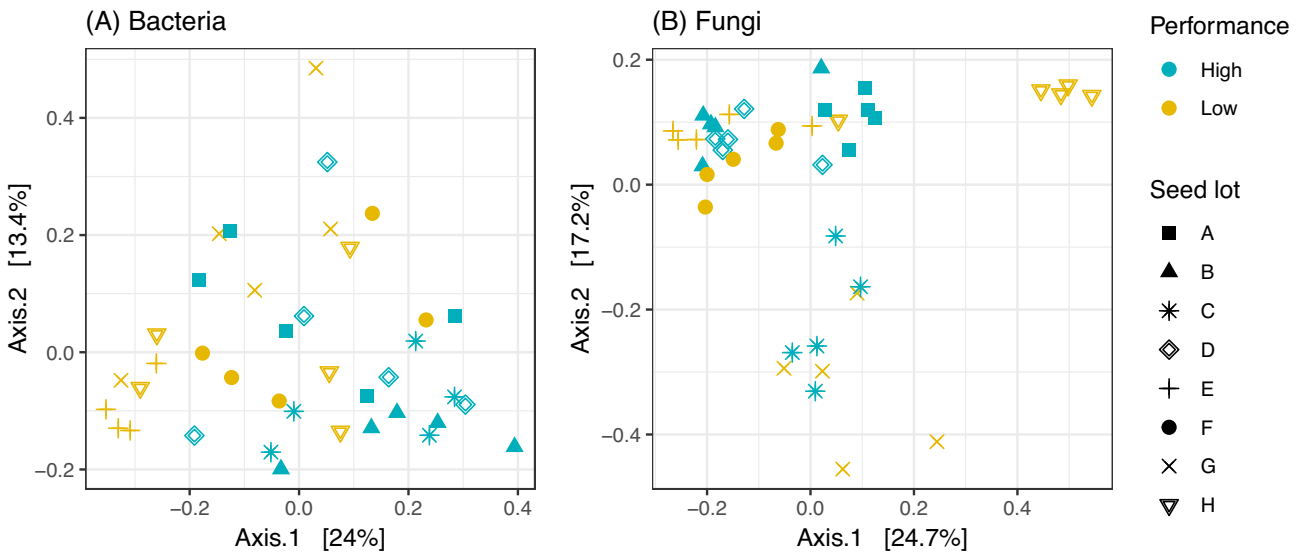

Figure 5. Bacterial (A) and fungal (B) community composition dissimilarity for spinach seed microbiota. Ordination plots based on non-metric multi-dimensional scaling (NMDS), using the Bray–Curtis distance metric. Samples of the eight seed lots are depicted using different shapes, while the colours blue and yellow are used to depict which seed lots are high- and low-performing, respectively.

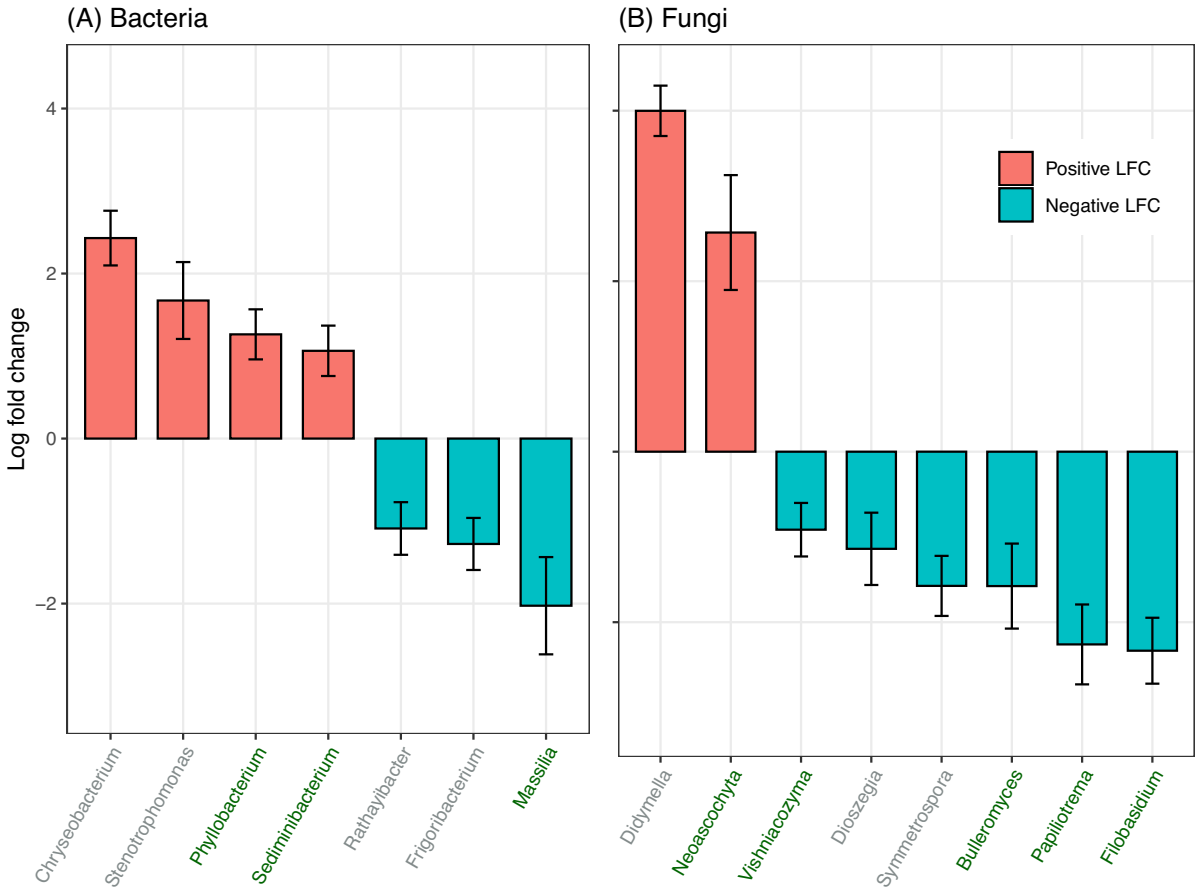

Figure 6. Differentially abundant bacterial (A) and fungal (B) genera. The depicted genera were identified as differentially abundant between the two seed lot performance groups (high/low performance). Results of the ANCOM-BC2 framework ($\alpha = 0.05$; P-values adjusted using the Benjamini–Hochberg procedure). The y-axis plots the (natural) log fold change (LFC), with negative values (blue bars) indicating that a genus is found in higher relative abundance in high-performing seed lots and positive values (red bars) indicating that a genus is found in higher relative abundance in seed lots of low performance. The colour-coding of the genera names indicates whether a genus successfully passed the sensitivity analysis for pseudo-count addition (bold green) or not (grey).

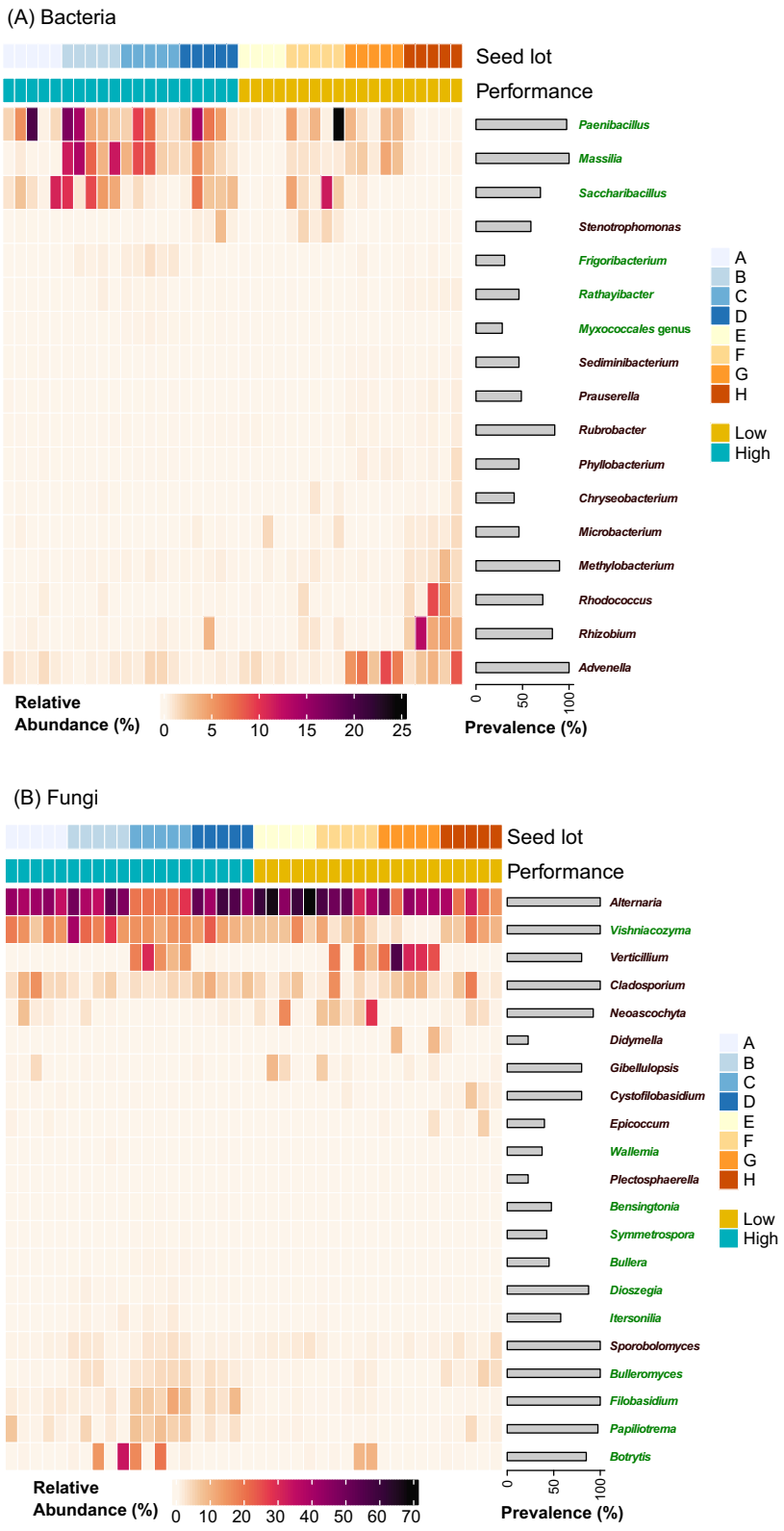

Figure 7. Heatmap of differentially abundant bacterial (A) and fungal (B) genera. The genera were identified as differentially abundant between the two seed lot performance groups (high/low performance), either by the ANCOM-BC2 framework, a Wald test using the DESeq2 R package or a Wilcoxon rank-sum test using the metacoder R package ($\alpha = 0.05$; *P*-values adjusted using the Benjamini–Hochberg procedure). Each vertical line of heat cells represents a different sample, with seed lot and seed lot performance group presented with different colour-codes at the top of the graph. The colour of the heat cells represents relative abundance (%) per genus. A grey bar per genus represents the percentage of samples (prevalence) that it is present in across all samples. The names of genera that are present in higher relative abundance in the high performance group are coloured green.

Vishniacozyma, *Mycosphaerella*, *Stemphylium*, *Cladosporium*, *Holtermanniella*, *Neoascochyta*, *Filobasidium*, *Papiliotrema*, *Bulleromyces*, *Botrytis*, *Sporobolomyces*, and *Dioszegia* (Supplementary File S2: Fig. S5).

Seed microbiota composition differs per seed lot and level of disease suppression

Out of the total variance in bacterial communities, 26.7% of variance is explained by the factor seed lot and 7.6% by the factor disease suppression when using the Jaccard distance (NMDS; PERMANOVA $\alpha = 0.05$). These proportions were 30.5% and 9.8%, respectively, when using the Bray–Curtis distance (NMDS; PERMANOVA $\alpha = 0.05$) (Fig. 5). Out of the total variance in fungal communities, 41.4% of variance is explained by the factor seed lot and 6.2% by the factor disease suppression (NMDS; PERMANOVA $\alpha = 0.05$). These proportions were 50.7% and 7.1% respectively when using the Bray–Curtis distance (NMDS; PERMANOVA $\alpha = 0.05$) (Fig. 5).

Differentially abundant taxa across performance groups

Detailed information on results presented in this section are documented in Supplementary File S5 for all three statistical methods used, namely the ANCOM-BC2 framework, a Wald test using the DESeq2 R package and a Wilcoxon rank-sum test using the metacoder R package. The results of the ANCOM-BC2 framework are presented in Fig. 6 for bacterial and fungal genera, while the results of metacoder and DESeq2 can be found in the supplementary information (metacoder–Supplementary File S2: Fig. S6; DESeq2–Supplementary File S2: Fig. S7). The heatmaps of Fig. 7 include all bacterial and fungal genera that were identified as differentially abundant by any of the three methods used.

The bacterial genera, *Massilia* (2.49%), *Saccharibacillus* (2.01%), and a *Myxococcales* genus (0.02%), were found in higher relative abundance in high-performing seed lots as indicated by two of the three methods used while *Advenella* (1.56%), *Phyllobacterium* (0.08%), *Chryseobacterium* (0.07%), and *Sediminibacterium* (0.02%) were found in higher relative abundance in low-performing seed lots by two of the three methods used (Table 1). The fungal genera, *Vishniacozyma* (11.40%), *Filobasidium* (1.64%), *Papiliotrema* (1.54%), *Itersonilia* (0.15%), *Dioszegia* (0.10%), *Bullera* (0.06%), *Wallemia* (0.02%), *Symmetrospora* (0.02%), and *Bensingtonia* (0.01%) were identified as more abundant in high-performing seed lots, by at least two of the three methods, while the genera *Neoascochyta* (2.34%), *Didymella* (0.47%), and *Plectosphaerella* (0.01%) were found in higher relative abundance in low-performing seed lots by two of the three methods used (Table 1). Differentially abundant bacterial and fungal taxa identified at the species level are presented in Supplementary File S5. Finally, genera that were either only present in high- or only present in low-performing seed lots are presented in Supplementary File S2: Table S1, with most of these taxa being of low relative abundance and low prevalence.

Indicator taxa per performance group

A set of bacterial and fungal genera were identified as indicators of high- and low-performing seed lots (BH; $\alpha = 0.05$). For bacteria, *Massilia*, *Frigoribacterium*, and a *Myxococcales* genus were indicators associated with high-performing seed lots, while *Advenella*, *Marmoricola*, *Rubrobacter*, *Brevundimonas*, *Microbacterium*, *Nocardioides*, *Sediminibacterium*, *Xanthomonas*, *Phyllobacterium*, and *Rhodococcus* were associated with low-performing seed lots (Supplementary File S6). For fungi, *Vishniacozyma*, *Filobasidium*, *Papiliotrema*, *Dioszegia*, *Itersonilia*, *Bullera*, *Wallemia*, *Bensingtonia*, and *Symmetrospora* were indicators associated with high-performing seed lots, while *Didymella* with low-performing seed lots (Supplementary File S6).

Discussion

The composition of the spinach seed microbiota

In agreement with our hypothesis, we confirmed that part of the variance in microbial community composition was explained by the factor seed lot. This is in line with multiple studies highlighting the importance of plant genotype (Chen et al. 2020, Davies et al. 2024, Morales Moreira et al. 2021a), as well as terroir (Klaedtke et al. 2016, Morales Moreira et al. 2021b) and seed processing (Abdelfattah et al. 2023) in shaping the seed microbiota. Apart from exploring these differences, we used our dataset to describe core spinach seed microbiota and explore the ubiquity of rare taxa in our samples. To our knowledge, there is, at present, only one other study specifically focusing on spinach seed microbiota. Kandel et al. (2022) used DNA samples recovered from the pericarp of spinach seeds to report on the taxonomical composition of seed epiphytes using different seed lots. Similarly to their findings, our results reinforce the prevalence of *Pantoea*, *Pseudomonas*, *Curtobacterium*, *Massilia*, *Sphingomonas*, and *Sanguibacter* as core bacterial taxa and of *Alternaria* and *Botrytis* as core fungal taxa of spinach seed microbiota (Fig. 2; Supplementary File S2: Fig. S5). The thorough DNA extraction method we developed, allowed us to extract both epiphytic and endophytic microbial DNA from our samples. This may be the reason why, unlike Kandel et al. (2022), we describe numerous yeast genera as core fungal taxa of spinach seed microbiota, some of which may reside in the seeds endophytically. We used and recommend the same DNA extraction method for other plant species. Specifically, we verified its suitability for extracting (bacterial) DNA from red fescue, perennial ryegrass and beetroot seeds (Supplementary File S7).

On a broader level, similarly to our findings in spinach, the Seed Microbiota Database (Simonin et al. 2022) reports *Pantoea*, *Pseudomonas*, *Paenibacillus*, *Sphingomonas*, *Rhizobium*, and *Methylobacterium* (bacterial genera) as well as *Alternaria*, *Vishniacozyma*, *Cladosporium*, *Filobasidium*, and *Sporobolomycetes* (fungal genera) as core seed microbiota across multiple plant hosts. Interestingly, Simonin et al. (2022) highlight the prevalence of *Alternaria metachromatica*, which was also the most prevalent fungal species in our study.

Although seed microbiome studies do not distinguish between yeasts and filamentous fungi, we noticed the prevalence of multiple yeast genera. We expect that the presence of yeasts in seeds is facilitated by their ubiquity in the phyllosphere and more specifically the anthosphere, where seeds develop (Aleklett et al. 2014, Kemler et al. 2017).

Apart from the identity of core microbiota and the distinct presence of yeast taxa, all samples included numerous rare taxa. Ecosystems of all dimensions are characterized by the dichotomy of having few dominant (core) and a multitude of rare taxa. While rarity is linked to stochastic events, rare taxa are thought to be crucial for stable ecosystem functioning based on the insurance theory of biodiversity, especially in the presence of a stressor (Jousset et al. 2017, Van Nes et al. 2024). Similarly to many other ecosystems, seed microbiomes are also largely shaped by rare taxa (Johnston-Monje et al. 2022) and there is evidence to suggest that such taxa confer community robustness in seed microbiota as well (Kim et al. 2023). We calculated that an overwhelming 81% of the bacterial and 79% of the fungal taxa of our study were rare.

Table 1. Bacterial (A) and fungal (B) genera identified as differentially abundant. The depicted genera were identified as differentially abundant between the two seed lot performance groups (high/low performance), using three statistical methods: the ANCOM-BC2 framework (ANCOMBC R package), a Wald test (DESeq2 R package) or a Wilcoxon rank-sum test (metacoder R package) ($\alpha = 0.05$; *P*-values adjusted using the Benjamini–Hochberg procedure). Genera were present in higher relative abundance either in high- or low-performing seed lots. An asterisk (*) designates which genera are part of the core microbiota and genera in bold were reported by more than one method. The table also includes relative abundance (percentage of reads per genus per performance group) and prevalence (percentage of samples where a genus is found per performance group).

A. Bacteria	ANCOM-BC2	DESeq2	Metacoder	Abundance % in high performance group	Abundance % in low performance group	Prevalence % in high performance group	Prevalence % in low performance group
Present in higher relative abundance in high-performing seed lots							
*Paenibacillus**			x	5.88	2.4	100	95
Massilia*	x	x		4.22	0.81	100	100
Saccharibacillus		x	x	3.02	1.06	95	42
Frigoribacterium	x			0.2	0.02	45	16
Rathayibacter	x			0.06[a]	0.08	45	47
Myxococcales genus		x	x	0.04	0.001	50	5
Present in higher relative abundance in low-performing seed lots							
Advenella*		x	x	0.54	2.72	100	100
*Rhizobium**		x		0.26	1.54	80	84
Stenotrophomonas	x			0.21	0.24	65	53
*Methylobacterium**		x		0.11	0.44	85	95
Rhodococcus		x		0.05	0.98	60	84
Microbacterium		x		0.04	0.32	35	58
*Rubrobacter**		x		0.02	0.1	80	89
Phyllobacterium	x	x		0.02	0.14	40	53
Chryseobacterium	x	x		0.02	0.12	35	47
Prauserella		x		0.01	0.08	40	58
Sediminibacterium	x	x		0.01	0.04	40	53

B. Fungi	ANCOM-BC2	DESeq2	Metacoder	Abundance % in high performance group	Abundance % in low performance group	Prevalence % in high performance group	Prevalence % in low performance group
Present in higher relative abundance in high-performing seed lots							
Vishniacozyma*	x		x	16.24	6.56	100	100
*Botrytis**		x		4.26	1	80	90
Filobasidium*	x	x	x	3.08	0.19	100	100
Papiliotrema*	x	x	x	2.69	0.39	100	95
*Bulleromyces**	x			1.59	0.79	100	100
Itersonilia		x	x	0.27	0.02	85	30
Dioszegia*	x	x	x	0.17	0.03	100	75
Bullera		x	x	0.12	0.003	80	10
Wallemia		x	x	0.03	0.0004	65	10
Symmetrospora	x		x	0.03	0.002	60	25
Bensingtonia		x	x	0.02	0.002	70	25
Present in higher relative abundance in low-performing seed lots							
*Alternaria**		x		38.98	42.04	100	100
*Cladosporium**		x		4.69	4.83	100	100
Verticillium		x		4.27	10.22	70	90
*Sporobolomyces**		x		1.02[a]	0.96	100	100
Neoascochyta*	x	x		0.79	3.88	100	85
Gibellulopsis		x		0.26	1.1	75	85
Epicoccum		x		0.05	0.44	45	35
Cystofilobasidium		x		0.05	0.6	80	80
Didymella	x	x		0.001	0.93	10	35
Plectosphaerella		x	x	0.001	0.02	5	40

[a]Note that the log-fold change generated by ANCOM-BC2, represents the difference in bias-corrected abundances between groups and does not directly reflect relative abundance data. Similarly, log-fold change values generated by DESeq2 may not directly reflect relative abundance data due to the normalization procedure that is inherent to this method.

For fungi, species richness and Shannon diversity index positively correlated with disease suppression (Fig. 4; Supplementary File S2: Fig. S4), implying that an increase in fungal diversity may contribute to disease suppression against *G. ultimum*, possibly due to the occupation of more niches.

Yeasts and other seed microbiota are associated with disease suppression

We also hypothesized and confirmed that differences in taxonomical composition correlate with seed lot disease suppression. Level of disease suppression was described per seed lot based on the results of two spinach—*G. ultimum* bioassays. Together, the eight seed lots provided us with a stable gradient of disease suppression against *G. ultimum* damping-off and allowed us to explore the different seed microbiota for correlations between levels of disease suppression per seed lot and seed microbiome characteristics. We showed that 9.8% of bacterial and 7.1% of fungal community variance between samples correlated with disease suppression (Fig. 5), thus, pointing at the possible contribution of seed microbiota in disease suppression in this particular pathosystem. Seed lot performance also correlated with differences in relative abundance of certain taxa with multiple of those belonging to the core microbiota we described (Table 1).

The suppressive effect of bacterial taxa mostly related to the increased abundance of the genus *Massilia* in high-performing seed lots. This genus was also part of the core microbiota and found in all samples. Numerous studies have linked them to suppressiveness against plant pathogens (Andreo-Jimenez et al. 2021, Lv et al. 2024, Siegel-Hertz et al. 2018). With regards to fungi, biodiversity correlated positively with seed lot disease suppression while numerous taxa were also found in higher relative abundance in high-performing seed lots. It was particularly interesting to find *Vishniacozyma*, *Filobasidium*, *Papiliotrema*, and *Dioszegia* as differentially abundant taxa since these are also core taxa amongst the seed lots of our study. Together with the genera *Itersonilia*, *Bullera*, and *Bulleromyces* which are also more highly present in high-performing seed lots, the above genera belong to the *Tremellomycetes* class of dimorphic basidiomycetous fungi.

Dimorphic species can be found in either a yeast state or hyphal form. Dimorphism is often observed in species with multiple nutrient acquisition strategies, where the yeast state can be a resting structure or represent a saprobic asexual state, while the hyphal form associates with parasitism on either animals, plants or other fungi (Begerow et al. 2017). Among others, the genera *Papiliotrema*, *Dioszegia*, and *Bullera* are known for being able to parasitize other fungi (Begerow et al. 2017). Interestingly, although yeasts are mostly known as members of the Ascomycota, the majority of known and potentially mycoparasitic yeasts belong to the Basidiomycota phylum (Begerow et al. 2017).

Apart from mycoparasitism, there have been multiple other traits reported in yeasts, which makes them efficient biological control agents (BCAs). Different yeasts have been studied and used as BCAs, particularly against post-harvest pathogens given their ability to compete for space (niche occupation) and nutrients, secrete polymer-cleaving enzymes which damage the cell walls of pathogens, produce toxins, and volatile organic compounds (VOCs), induce host resistance and as previously mentioned parasitize other fungi (Freimoser et al. 2019). Our work indicates the possible importance of yeasts in controlling early stage seedling diseases such as *G. ultimum* damping-off.

From the *Vishniacozyma* genus, which we detected as most prevalent in high-performing seed lots, *V. victoriae* represents an example of successful biological control against post-harvest disease of fruits such as *Penicillium expansum, Botrytis cinerea*, and *Cladosporium* sp. (Gorordo et al. 2022, Nian et al. 2023). This species was also present in our dataset. The *Papiliotrema* genus also contains BCAs, such as the patented *P. terrestris* strain PT22AV, which is used against *Penicillium expansum* on apple post-harvest. This BCA was also antagonistic against *Botrytis cinerea, Rhizopus stolonifer, Aspergillus niger* and *Monilinia* spp. in other fruits (Ianiri et al. 2024). Additionally, *Papiliotrema flavescens* inhibits Fusarium-induced crown rot and head blight in wheat (Liu et al. 2021), while evidence suggests it can promote plant growth and induce systemic resistance via VOC production (Liu et al. 2024). Another mechanism through which the fungal fraction of the high-performing spinach seed lots mitigates *G. ultimum* disease could be niche occupation since suppressive seed lots have more diverse fungal communities. This would be in line with the fact that, apart from outcompeting the pathogen for space and resources, it has been shown that more diverse ecosystems are more capable to maintain ecosystem equilibrium in the presence of a stressor (Begerow et al. 2017, Kim et al. 2023, Van Nes et al. 2024).

The genera *Vishniacozyma*, *Filobasidium*, *Papiliotrema*, *Bulleromyces*, *Dioszegia*, *Itersonilia*, *Bensingtonia*, and *Bullera* are all common phylloplane yeasts found in plants of temperate climates, with known adaptations to that habitat such as protective pigment formation against high radiation (Buzzini et al. 2017). The book that reports this information covers the topic of yeast diversity in different natural ecosystems, such as the phylloplane, soil, and decomposing plant material. However, seed-associated yeasts are not mentioned (Buzzini et al. 2017). At the same time, seed microbiome studies seem to make no distinction between yeasts and filamentous fungi found in their samples. Our results indicate that this may be an important research gap and we recommend for studies to be shifted towards the presence and function of yeasts in seed microbiomes.

Temporal dynamics of disease suppression and relevance of seed microbiota

Apart from seed germination, the seedling stage is also critical for plant health. Seedling microbiota assembly is a dynamic process in which seed- and soil-derived microorganisms compete for a share of carbon-rich seed exudates and for the possibility of securing a niche inside or in close proximity to the new-born plant (Barret et al. 2015, Torres-Cortés et al. 2018). Although certain seed-derived taxa remain part of the microbiota of the emerging seedling, multiple studies suggest that seedling microbiota are predominantly recruited from the soil (Escobar Rodríguez et al. 2020, Ofek et al. 2011, Rochefort et al. 2021). It would be coherent to infer that the role of the seed microbiome in plant health is more pronounced during germination, when external biotic conditions cannot be as influential as later. While the experimental setup of our study does not allow the validation of this, we hypothesize that the yeasts we described are able to antagonize *G. ultimum* during germination when seed exudation makes sugars readily available, due to their known antagonistic effects and ability to multiply fast. This is particularly important to mitigating infection by *G. ultimum* since this is a very fast pathogen, with its sporangia being reported to respond to the release of seed exudates within 30 min after exposure (Windstam and Nelson 2008). Lastly, since the zoospores of phytopathogenic oomycetes are known to use seed and root exudates as chemical cues guid-

ing them towards the host plant, the uptake of such compounds by seed-associated yeasts possibly masks seed germination (Deacon 1996, Kasteel et al. 2023).

Multiple studies indicate that seed microbiota comprise aboveground plant-associated or airborne taxa which colonize seeds during flowering and seed maturation (Chesneau et al. 2020, Rezki et al. 2018, Torres-Cortés et al. 2018). We observe this to be the case for most of the taxa in our study that were associated with high-performing seed lots. Therefore, since these organisms are not adapted to living below-ground, we speculate that after an initial surge in numbers, they are outcompeted and die off, while young seedlings are colonized by neighbouring soil-dwelling microorganisms (Escobar Rodríguez et al. 2020, Rochefort et al. 2021). This is why we speculate that seed-associated yeasts are of particular importance specifically for disease suppression against fast pathogens when the onset of germination is a crucial timepoint. We also expect pathogens using chemotaxis to be more vulnerable to this means of microbiota-induced disease suppression. This is also in line with our previous study, where seed microbiota were able to confer disease suppression against *G. ultimum* but not against pathogens attacking seedlings at a later stage or against pathogens that do not utilize chemical cues during pathogenesis (Diakaki et al. 2022).

Utilizing indicator taxa to ensure seedling health

The use of pathogen-free seeds is the first step in ensuring plant health. This is the main reason why seed producing companies occasionally need to disinfect seeds prior to commercialization. Yet seed disinfection may inadvertently affect the seed microbiota which, as we have demonstrated, may be an asset for seed and seedling health during the onset of germination. We consider our results to be a preliminary source of information for identifying indicator taxa, which could be used for making informed decisions on seed disinfection. Upon validation of our results, spinach seed microbiota could be screened for the relative abundance of certain taxa as a more efficient method of estimating vulnerability to *Globisporangium* damping-off. When testing this further, we recommend focusing on the core taxa that are differentially abundant between high and low-performing seed lots, namely *Massilia*, *Vishniacozyma*, *Filobasidium*, and *Papiliotrema* as potential indicators of suppressive and *Neoascochyta* as indicator of non-suppressive seed lots.

Conclusions

We were able to confirm our hypothesis and prove that the seed microbiota of the eight spinach seed lots of our study differed in taxonomic composition, while part of these differences correlated with suppressiveness against *G. ultimum* damping-off. Our results point at the ubiquity of basidiomycetous dimorphic yeasts in spinach seed microbiota. We especially highlight the possible importance of these microorganisms in mitigating infection by pathogens attacking plants during germination such us the oomycete *G. ultimum* even though they are not adapted to below-ground conditions. While many studies only focus on bacteria, our findings support the importance of including fungi in studies exploring the benefits of seed microbiota in plant health. Finally, upon validation of these preliminary results, we consider our findings to be applicable in the seed industry by pointing at potential indicator taxa for making informed decisions on seed disinfection.

Acknowledgements

We thank Els Nijhuis for facilitating the processing of the data with Qiime2 and Dr Sara Cazzaniga for providing guidance for the implementation of the ANCOM-BC2 framework. We thank Bejo Zaden B.V., Nunhems Netherlands B.V. (BASF), Pop Vriend Seeds B.V., and Sakata Vegetables Europe S.A.S. for providing spinach seed lots. The project 'Changing the system of seed health; An initiative of industry and research towards a paradigm shift' (LWV19097) receives financial support from the Top Sector Horticulture and Propagation Materials. Within the Top Sector, the business community, knowledge institutions, and the government work together on innovations in the field of sustainable production of healthy and safe food and the development of a healthy, green living environment.

Author contributions

Makrina Diakaki (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft), Beatriz Andreo Jimenez (Formal analysis, Visualization, Writing – review & editing), Ezra de Lange (Investigation, Methodology, Writing – review & editing), Patrick Butterbach (Investigation, Methodology, Validation), Liesbeth van der Heijden (Investigation, Methodology, Validation, Writing – review & editing), Jürgen Köhl (Conceptualization, Funding acquisition, Writing – review & editing), Wietse de Boer (Conceptualization, Supervision, Writing – review & editing), and Joeke Postma (Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing)

Supplementary data

Supplementary data is available at *FEMSEC Journal* online.

Conflict of interest: Patrick Butterbach is employed at Germains Seed Technology and Liesbeth van der Heijden at Bejo Zaden B.V.

Funding

This work is part of the project 'Changing the system of seed health; An initiative of industry and research towards a paradigm shift' (LWV19097) which receives financial support from the Top Sector Horticulture and Propagation Materials. The project is funded by the Dutch Ministry of Agriculture, Nature and Food Quality, Bejo Zaden B.V., CN Seeds Ltd, DLF B.V., e-nema Gesellschaft für Biotechnologie und biologischen Pflanzenschutz mbH, Germains Seed Technology, Nunhems Netherlands B.V. (BASF), Pop Vriend Seeds B.V., and Sakata Vegetables Europe S.A.S.

References

Abdelfattah A, Tack AJ, Lobato C *et al.* From seed to seed: the role of microbial inheritance in the assembly of the plant microbiome. *Trends Microbiol* 2023;**31**:346–55.

Aleklett K, Hart M, Shade A. The microbial ecology of flowers: an emerging frontier in phyllosphere research. *Botany* 2014;**92**:253.

Andreo-Jimenez B, Schilder MT, Nijhuis EH *et al.* Chitin-and keratin-rich soil amendments suppress *rhizoctonia solani* disease via changes to the soil microbial community. *Appl Environ Microb* 2021;**87**:11.

Apprill A, Mcnally S, Parsons R *et al.* Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* 2015;**75**:129.

Barret M, Briand M, Bonneau S *et al.* Emergence shapes the structure of the seed microbiota. *Appl Environ Microb* 2015;**81**:1257.
Begerow D, Kemler M, Feige A *et al.* Parasitism in yeasts. In: Buzzini P, Lachance MA, Yurkov A (eds), *Yeasts in Natural Ecosystems: Ecology*. Cham: Springer, 2017, 179–210.
Bewley JD, Bradford KJ, Hilhorst HWM *et al.* Germination. In: Bewley JD, Bradford KJ, Hilhorst HWM, *et al.* (eds), *Seeds: Physiology of Development Germination and Dormancy*. New York: Springer, 2013, 133–81.
Bisanz JE. qiime2R: importing QIIME2 artifacts and associated data into R sessions. 2018.
Bokulich NA, Kaehler BD, Rideout JR *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 2018;**6**:90.
Bolyen E, Rideout JR, Dillon MR *et al.* Reproducible interactive scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;**37**:8.
Buzzini P, Lachance MA, Yurkov A. *Yeasts in Natural Ecosystems: diversity*. Cham: Springer, 2017.
Callahan BJ, McMurdie PJ, Rosen MJ *et al.* DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;**13**:581.
Chen X, Krug L, Yang H *et al. Nicotiana tabacum* seed endophytic communities share a common core structure and genotype-specific signatures in diverging cultivars. *Comput Struct Biotechnol J* 2020;**18**:287–95.
Chesneau G, Torres-Cortes G, Briand M *et al.* Temporal dynamics of bacterial communities during seed development and maturation. *FEMS Microbiol Ecol* 2020;**96**:12.
Davies J, Hawkins S, Winters A *et al.* Bacterial endophytic community composition varies by hemp cultivar in commercially sourced seed. *Environ Microbiol Rep* 2024;**16**:2.
Deacon JW. Ecological implications of recognition events in the pre-infection stages of root pathogens. *New Phytol* 1996;**133**:135.
De Cáceres M, Legendre P. Associations between species and groups of sites: indices and statistical inference. *Ecology* 2009;**90**:3566.
Diakaki M, van der Heijden L, Lopez-Reyes JG *et al.* Beetroot and spinach seed microbiomes can suppress *pythium ultimum* infection: results from a large-scale screening. *Seed Sci Res* 2022;**32**:274.
Escobar Rodríguez C, Antonielli L, Mitter B *et al.* Heritability and functional importance of the *Setaria viridis* bacterial seed microbiome. *Phytobiomes J* 2020;**4**:1.
Foster Z, Sharpton T, Grünwald N. Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. *PLOS Comput Biol* 2017;**13**:e1005404.
Freimoser FM, Rueda-Mejia MP, Tilocca B *et al.* Biocontrol yeasts: mechanisms and applications. *World J Microbiol Biotechnol* 2019;**35**:10.
Gardes M, Bruns TD. ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. *Mol Ecol* 1993;**2**:113.
Gorordo MF, Lucca ME, Sangorrín MP. Biocontrol efficacy of the *Vishniacozyma victoriae* in semi-commercial assays for the control of postharvest fungal diseases of organic pears. *Curr Microbiol* 2022;**79**:9.
Hartig F. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models R Package Version 0.2. 4. 2019.
Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. *Biometrical J* 2008;**50**:346.
Husenov B, Asaad S, Muminjanov H *et al.* Sustainable wheat production and food security of domestic wheat in Tajikistan: implications of seed health and protein quality. *Int J Environ Res Public Health* 2021;**18**:5751.
Ianiri G, Barone G, Palmieri D *et al.* Transcriptomic investigation of the interaction between a biocontrol yeast *Papiliotrema terrestris* strain PT22AV and the postharvest fungal pathogen *penicillium expansum* on apple. *Commun Biol* 2024;**7**:1.
ISTA. *International Rules for Seed Testing*. Bassersdorf: International Seed Testing Association. 2020.
Jack ALH, Nelson EB. A seed-recruited microbiome protects developing seedlings from disease by altering homing responses of *Pythium aphanidermatum* zoospores. *Plant Soil* 2018;**422**:209–22.
Johnston-Monje D, Gutiérrez JP, Becerra Lopez-Lavalle LA. Seed-transmitted bacteria and fungi dominate juvenile plant microbiomes. *Front Microbiol* 2021;**12**:737616.
Johnston-Monje D, Gutiérrez JP, Becerra Lopez-Lavalle LA. Stochastic inoculum biotic filtering and species-specific seed transmission shape the rare microbiome of plants. *Life* 2022;**12**:1372.
Jousset A, Bienhold C, Chatzinotas A *et al.* Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J* 2017;**11**:853.
Kandel SL, Henry PM, Goldman PH *et al.* Composition of the microbiomes from spinach seeds infested or noninfested with *peronospora effusa* or *verticillium dahliae*. *Phytobiomes Journal* 2022;**6**:169.
Kasteel M, Ketelaar T, Govers F. Fatal attraction: how *phytophthora* zoospores find their host. *Semin Cell Dev Biol* 2023;**148**:13–21.
Kemler M, Witfeld F, Begerow D *et al.* Phylloplane yeasts in temperate climates. In: Buzzini P, Lachance MA, Yurkov A (eds), *Yeasts in Natural Ecosystems: diversity*. Cham: Springer, 2017, 171–97.
Kim H, Kim C, Lee YH. The single-seed microbiota reveals rare taxa-associated community robustness. *Phytobiomes J* 2023;**7**:3.
Klaedtke S, Jacques MA, Raggi L *et al.* Terroir is a key driver of seed-associated microbial assemblages. *Environ Microbiol* 2016;**18**:1792.
Köhl J, Groenenboom-de Haas L, Houwers I *et al.* Generic viability qPCR for monitoring shelf life of microbial biological control agents coated on seeds based on the nucleic acid intercalating dyes EMA and PMAxx. *Biol Control* 2024;**196**:105569.
Lahti L, Shetty S. Introduction to the microbiome R package. 2018.
Lin H, Eggesbø M, Peddada SD. Linear and nonlinear correlation estimators unveil undescribed taxa interactions in microbiome data. *Nat Commun* 2022;**13**:1.
Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. *Nat Commun* 2020;**11**:1.
Lin H, Peddada SD. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. *Nat Methods* 2024;**21**:83.
Liu S, Xie J, Luan W *et al. Papiliotrema flavescens* a plant growth-promoting fungus alters root system architecture and induces systemic resistance through its volatile organic compounds in *Arabidopsis*. *Plant Physiol Biochem* 2024;**208**:108474.
Liu Z, Li X, Sun Z *et al. Papiliotrema flavescens* colonized in biochars inhibits wheat crown rot and *fusarium* head blight. *Biochar* 2021;**3**:625.
Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:1–21.
Lundberg DS, Yourstone S, Mieczkowski P *et al.* Practical innovations for high-throughput amplicon sequencing. *Nature* 2013;**10**:10.
Lv N, Ravanbakhsh M, Ling S *et al.* Harnessing key bacteria from suppressive soil to mitigate banana Panama disease. *Biol Fertil Soils* 2024;**60**:845–52.
Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j* 2011;**17**:10.
Matsumoto H, Fan X, Wang Y *et al.* Bacterial seed endophyte shapes disease resistance in rice. *Nat Plants* 2021;**7**:60.

McGuire S, Sperling L. Les liens entre la sécurité alimentaire et la sécurité des semences: faits et fiction qui orientent la riposte. *Dev Pract* 2011;**21**:4–5.
McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;**8**:e61217.
Mesterházy Á, Oláh J, Popp J. Losses in the grain supply chain: causes and solutions. *Sustainability* 2020;**12**:2342.
Morales Moreira ZP, Helgason BL, Germida JJ. Environment has a stronger effect than host plant genotype in shaping spring *brassica napus* seed microbiomes. *Phytobiom J* 2021;**5**:220.
Morales Moreira ZP, Helgason BL, Germida JJ. Crop genotype and field environmental conditions shape bacterial and fungal seed epiphytic microbiomes. *Can J Microbiol* 2021;**67**:161.
Nearing JT, Douglas GM, Hayes MG *et al.* Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun* 2022;**13**:1.
Nelson EB. The seed microbiome: origins interactions and impacts. *Plant Soil* 2018;**422**:1–5.
Nian L, Xie Y, Zhang H *et al. Vishniacozyma victoriae*: an endophytic antagonist yeast of kiwifruit with biocontrol effect to *Botrytis cinerea*. *Food Chem* 2023;**411**:135442.
Nilsson RH, Larsson KH, Taylor AFS *et al.* The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res* 2019;**47**: D259.
Ofek M, Hadar Y, Minz D. Colonization of cucumber seeds by bacteria during germination. *Environ Microbiol* 2011;**13**:2794.
Oksanen J. Constrained ordination: tutorial with R and vegan. *R-packace Vegan* 2012;**1**:10.
Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities time series and global field samples. *Environ Microbiol* 2016;**18**:1403.
Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2012;**41**:D590.
R Core Team. *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2024.
Rai M, Abd-Elsalam KA, Ingle AP. *Pythium Diagnosis Diseases and Management*. Boca Raton: CRC Press, 2020.
Rezki S, Campion C, Simoneau P *et al.* Assembly of seed-associated microbial communities within and across successive plant generations. *Plant Soil* 2018;**422**:67–79.
Rochefort A, Simonin M, Marais C *et al.* Transmission of seed and soil microbiota to seedling. *Msystems* 2021;**6**:3.
Siegel-Hertz K, Edel-Hermann V, Chapelle E *et al.* Comparative microbiome analysis of a fusarium wilt suppressive soil and a fusarium wilt conducive soil from the Châteaurenard region. *Front Microbiol* 2018;**9**:568.
Simonin M, Briand M, Chesneau G *et al.* Seed microbiota revealed by a large-scale meta-analysis including 50 plant species. *New Phytol* 2022;**234**:1448.
Torres-Cortés G, Bonneau S, Bouchez O *et al.* Functional microbial features driving community assembly during seed germination and emergence. *Front Plant Sci* 2018;**9**:902.
Van Nes EH, Pujoni DGF, Shetty SA *et al.* A tiny fraction of all species forms most of nature: rarity as a sticky state. *P Natl Acad Sci USA* 2024;**121**:2.
Verma KS, White JF. *Seed Endophytes*. Cham: Springer, 2019.
White TJ, Bruns TD, Lee S *et al.* Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gefland DH, Sninsky JJ *et al.* (eds), *PCR Protocols: a Guide to Method and Applications*. San Diego: Academic Press, 1990, 315–22.
Wickham H. *ggplot2: elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016.
Windstam S, Nelson EB. Differential interference with *pythium ultimum* sporangial activation and germination by *Enterobacter cloacae* in the corn and cucumber spermospheres. *Appl Environ Microb* 2008;**74**:4285.
Wolfgang A, Zachow C, Müller H *et al.* Understanding the impact of cultivar seed origin and substrate on bacterial diversity of the sugar beet rhizosphere and suppression of soil-borne pathogens. *Front Plant Sci* 2020;**11**:560869.

Received 18 October 2024; **revised** 23 December 2024; **accepted** 7 January 2025