University of Dundee

USING MICROBIOME COMPOSITION AS A BIOMARKER FOR CROP PRODUCTIVITY

Muhammad Amanullah Jilani - 180017551

BSc (Hons) Molecular Genetics
School of Life Sciences, The University of Dundee
8th March 2021
mjilani@dundee.ac.uk

The James Hutton Institute

I declare that the content of this project report is my own work and has not previously been submitted for any other assessment. The report is written in my own words and conforms to the University of Dundee's Policy on plagiarism and academic dishonesty. Unless otherwise indicated, I have consulted all of the references cited in this report.

_____

Muhammad Amanullah Jilani

Table of contents

List of Figures

List of Tables

Abstract

The interface between plant roots and soil hosts complex microbial communities collectively referred to as the plant microbiome. Similar to the microbiome inhabiting our digestive tract, microbes thriving at the root-soil interface can influence the plant's growth, health, development and ultimately crop productivity. In this project, I hypothesized that yield differences observed between two spinach fields is determined, at least partly, by differences in their below-ground microbiome. To test this hypothesis, I utilized an existing 16S rRNA gene sequencing library for spinach, and I investigated its composition using ecological indexes and multivariate statistical analysis done in R. My work indicates that the spinach microbiome is a 'gated community', in the sense that the number and distribution of bacteria in the vicinity of and on spinach roots is significantly different from unplanted soil communities (Kruskal-Wallis test, *p-value* < 0.05). Interestingly, the 'microhabitat effect', which is the effect that bacteria are associated with a given plant organ, exceeded the impact of the field on the composition of the spinach microbiome (Adonis, R2 = 0.7795; R2 = 0.3109; *p-value* < *0.05*, 5,000 permutations). Yet, I was able to identify a limited, but significant, number of bacteria whose abundance discriminated between more and less productive fields (DESeq2, *p-value* < 0.05, FDR corrected). Of note, some of these bacteria were associated with taxa previously reported as plant-beneficial bacteria, including nannocystaceae, oxalobacteraceae, rhizobiaceae and nocardiodaceae. These bacteria can be then further investigated for their functionality in silico and represent an ideal target to develop direct or indirect inoculants for agriculture.

1. Introduction

Plants comprise a vital component of the global food supply and as the world's population grows so does our need for faster and more sustainable food production[1]. To address this need, there have been several techniques proposed such as vertical farming[2], hydroponics[3], advanced plant breeding[4] and, more recently, microbiome exploitation[5].

The plant microbiome refers to the microbial communities, encompassing bacteria, fungi, oomycetes and other microorganisms living in association with and interacting with plant organs[6]. In particular, the microbiome at the root-soil interface is analogous to the microbiome of our digestive tract in the sense that it is vital for the growth of the plant, the uptake of minerals, and the prevention of disease[5]. As shown in the past, the presence or absence of certain bacteria in the below-ground microbiome of plants has been correlated with increased crop yields[7].

The differing areas that microbiomes reside in are known as 'microhabitats'. One of the most important plant microhabitats is called the rhizosphere; the rhizosphere is the thin, root-soil interface found in the below-ground plant microbiome[6]. It is known to host bacterial families responsible for enhancing mineral uptake and/or protecting plant pathogens. These bacterial families are known as plant-growth-promoting rhizobacteria (PGPRs)[8].

It is hypothesized that PGPRs can be used as "biofertilizers" or "biopesticides", depending on their mode of action[7]. While the real-world practicality of these microbes is still being fully understood, current data suggests that they are an effective way of increasing crop productivity if utilized correctly[9].

The rational exploitation of the plant microbiome for agriculture is of key importance for gaining a precise understanding of the molecular mechanisms regulating the assembly of the microbiome at the root-soil interface. Current models postulate a two-step selection

process for the belowground microbiome: plant-associated communities are derived from the surrounding soil biota, representing the main (although not the only) source of inoculum, and plant-mediated “checkpoints” modulate the composition of these communities in the rhizosphere and in the roots[10]. The molecular mechanisms underpinning these checkpoints rhizodeposition such as the secretion of photosynthesis-derived metabolites in the rhizosphere, the root system architecture (RSA), the way plants “explore” the soil and the plant immune system, ability to recognize and terminate microbial multiplication, have all been implicated in plant-microbiota interactions belowground[11]. Of note, all these processes display a host genetic component: differences have been identified between species and between genotypes of the same species[12].

The plant microbiome can be studied using a cultivation-dependent approach, which requires the isolation of microbial members of the community, or a cultivation-independent approach, where community composition is directly inferred from an evaluation of its molecular components (e.g., DNA) without microbial isolation[13]. Among the latter approach, the profiling of the 16S rRNA gene became a well-established technique to infer composition and phylogenetic relatedness of bacterial and archaeal communities in the environment[14]. Combined with high-throughput sequencing and bioinformatic analyses, this technique allows scientists to gather data from replicated experiments in a timely and cost-effective fashion and, consequently, allows for testing hypotheses implicating microbiome composition to plant performance[13].

The Circles EU project (https://circlesproject.eu/) is a 10-million-euro European Union (EU) funded project that aims to investigate the potential of natural microbiomes to be exploited for more sustainable production of a variety of crops, notably spinach and tomato. However, while information on the tomato microbiome and field performance are becoming

available[15], the spinach microbiome can be considered an “uncharted territory” which must first be characterized before utilization

To that extent, in this study, I explore the below-ground microbiome of spinach plants from two fields that differ in crop yield.

I aim to ascertain the effect of microbiome composition on crop yield through the ecological indexes Alpha and Beta diversity[16]. Then, I aim to identify bacterial families significantly differing between the two fields through computational analysis in R studio. Consequently, I also aim to identify bacterial families that constitute the core spinach microbiome as well as the core microbiome of each field and discuss bacterial families of potential interest for agricultural application within each microbiome composition.

2. Materials and Methods

This dissertation uses a previously generated 16S rRNA gene sequencing dataset for the spinach microbiota, part of the on-going, Horizon 2020-funded, Innovation Actions 'CIRCLES'. In this dissertation, I developed and applied a computational approach to analyse the existing microbiota sequencing information.

2.1 The 16S rRNA gene sequencing dataset for the spinach microbiota

The 16S rRNA bacterial library was generated by the Davide Bulgarelli lab at the University of Dundee at the James Hutton Institute.

Plants and soil specimens were sampled in the spring of 2019 in two farms, hereafter referred to as 'Dragoni' and 'Minzoni', located in the Emilia Romagna region (Italy). Soils of both fields display a comparable chemical composition and were subjected to the same treatments during spinach cultivation (Bulgarelli D., personal communication). However, spinach crops in the Dragoni farm yielded more than the ones in the Minzoni farm, 25 T/ha versus 17 T/ha, respectively (Bulgarelli D., personal communication). Of note, the two farms planted different spinach varieties: as varietal choice confounds with the field, in this dissertation I'll refer to a combinatorial "field effect" encompassing farm characteristics (including the choice of the varieties). For each field, ten sampling sites were identified, for each sampling site, ten contiguous plants were harvested, as such, a single sampling site corresponds to ten plants. This was because spinach plants were harvested early in the season, and consequently did not contain enough individual DNA to create a dataset of acceptable quality and quantity.

Next, for each – composite – sample, leaves were separated from the roots and rhizosphere. Additionally, ten samples of unplanted soil were collected at different points that were at a distance from any plant, this was done for both fields. DNA extraction was performed for the

roots and rhizosphere microhabitats, as well as the unplanted soil samples. The DNA preparation was subjected to selective PCR amplification of the 16S rRNA gene using primers 505F and 806R and subjected to Illumina MiSeq amplicon sequencing as previously described[8,9].

Two water samples were also taken from the water used to irrigate each field, but they were not retained for downstream analysis as explained in **Section 2.2.1**.

2.2 Microbiota data Analysis

The files containing the code used for data analysis have all been uploaded to GitHub, and are available at this link here (https://github.com/RealAmanJ/BSc_Dissertation). The code is split into multiple scripts and each contains a general title in the name as well as a header in the form of comments at the top of the code to inform the reader of the script's exact purpose. Any files used or mentioned that are not generated by scripts, are also available on the repository.

2.2.1 Data pre-processing

The steps taken to pre-process the codes are outlined in **Figure 1**.

First, the data was imported into R-Studio running R version 4.0.2 (2020-06-22) on a Windows 10 64-Bit (Build 19041) computer. Then, taxa whose sequence was classified as either 'chloroplast' or 'mitochondria', which likely represent host plant contaminants, were removed. Next, the dataset was filtered from sequences matching a list of common contaminants (available in the GitHub repository as: "*JH06_contaminant_ASVs_ids.txt*").

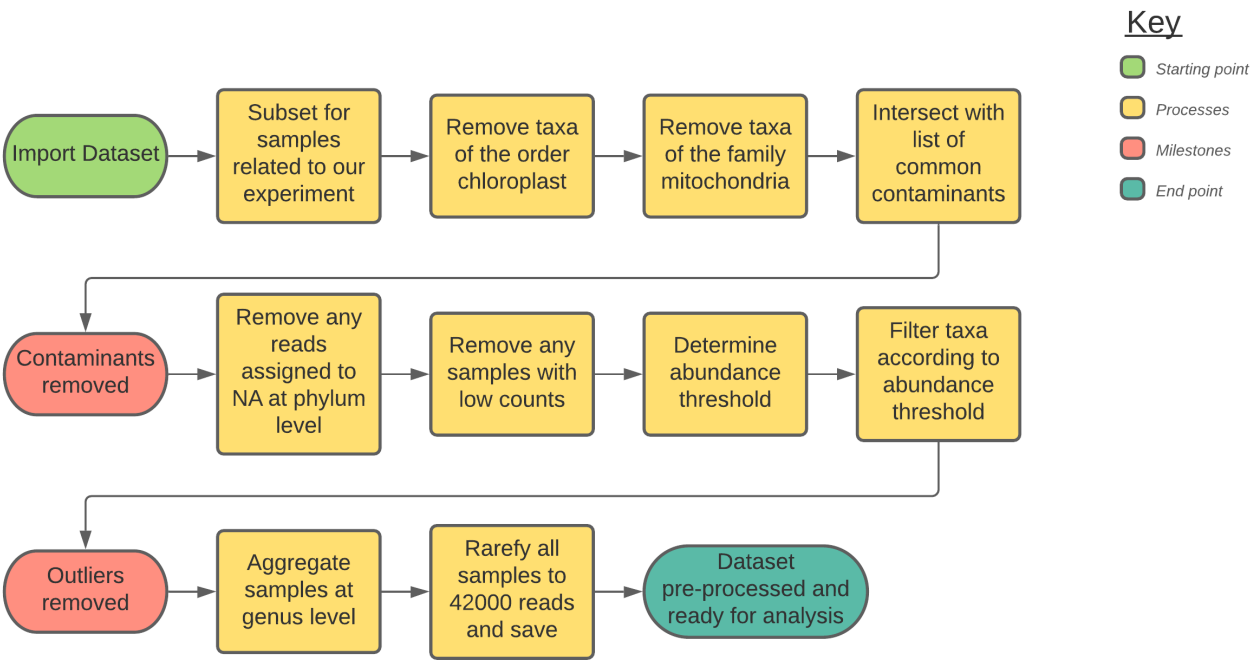


Figure 1: Pre-processing Flowchart
This figure illustrates the steps I took to process my data before the downstream analysis.

Afterwards, I removed any reads not matching entries at phylum level and any amplicon sequencing variants with low counts, likely representing a non-reproducible feature of the microbiome. The abundance threshold of 20 reads in at least 16% of the samples was determined according to previous investigations[15]. After having pruned these outliers in silico, I rarefied all samples to 42,000 reads and saved the dataset. It is noted that water samples – of which there were two – were lost, this is because a parallel analysis determined that the water samples have no identifiable impact and their presence only prevented the dataset from being rarefied at a higher resolution.

This dataset is from now on referred to as "*dat_info_42k*"

2.2.2 Alpha diversity

The 'estimate_richness' function of Phyloseq was used to create a new data frame (called "*dat_alpha_rare_42k*") with the measures defined as "Observed" and "Shannon".

Then, a new data frame (called "*design_alpha_42k*") was created by merging the sample data of *dat_info_42k* and the row names of "*dat_alpha_rare_42k"*.

Finally, columns of "*design_alpha_42k"* and "*dat_alpha_rare_42k"*, were bound with the cbind function to make dataset used for alpha diversity statistical analysis (called "*dat_alpha_rare_info_42k*").

The normality of distribution of the alpha diversity measures was then tested with the shapiro.test function.

2.2.3 Beta diversity

A bray-distance matrix was created by using the distance function of Phyloseq on *dat_info_42k* (called "*BC_42k*"). Following that, the ordinate command, again belonging to Phyloseq, was used on *dat_info_42k* with the method set as "CAP" and distance set to "Bray" (called as "*dat_info_CAP_BC_42k*"), and forms the basis of the Beta-diversity plot.

An analysis of variance (ANOVA) test was performed on *dat_info_CAP_BC_42k*, with the ANOVA function (permutations set to 5,000). An Adonis test was performed with *BC_42k* being the dependant variable and Microhabitat * Field the independent variable. The data used was the sample data of *dat_info_42k* and the permutations were again 5,000.

2.2.4 DESeq2 analysis and UpSetR plot preparation

The dataset *dat_info_42k* was split into two datasets corresponding to the Dragoni and Minzoni fields via the subset_samples function of Phyloseq (called "*dat_info_42k_dragoni*" and "*dat_info_42k_minzoni*", respectively).

The count data was extracted and then converted into a data frame, the sample data was also extracted and converted into a data frame (called "*countData*" and "*colData*" respectively). This was done for both *dat_info_42k_dragoni* and *dat_info_42k_minzoni*.

Afterwards, a DESeq dataset was created with the DESeqDataSetFromMatrix() command of DESeq2 from *countData* and *colData* with the design being set as microhabitat (called "*JH18_42k_cds*"). A differential count analysis was executed with the DESeq function of DESeq2 on *JH18_42k_cds* (called "*JH18_42k_cds_test*").

A contrast was then generated for two comparisons, Root vs Bulk and Rhizosphere vs Bulk, for each field (called "*JH18_42k_contrast*". This is visualized in **Supplementary Figure 1.** After generating the contrast, genera of adjusted p-value below 0.05 were extracted from *JH18_42k_contrast* and were then intersected with those corresponding to the Root and Rhizosphere microhabitats.

This procedure was done sequentially for both comparisons in both fields, each time, the list of differentially enriched bacteria corresponding to the microhabitat of interest was saved under a unique name.

Each one of the original four lists was then converted into a data frame. After conversion, a base mean column was added, and row names were renamed to have each bacterium identified correspond to its ASV read. The base mean columns were then converted to a Boolean format and lists of enriched root ASV reads, enriched rhizosphere ASV reads, and enriched spinach (core microhabitat) ASV reads, were created.

The enriched spinach ASV reads were then added to each of the four data frames, and then their row names were renamed to those of the same enriched spinach ASV list. These were all merged to create the spinach ASV data frame that forms the basis of the UpSetR plot (**Figure 4)**.

Lists of intersections and their corresponding Phyloseq objects were created by using the prune_taxa command of Phyloseq for each intersection list and *dat_info_42k* were saved.

2.3 Graphical outputs

2.3.1 UpsetR plot

All previous lists were merged into one and, with the upset command of the UpSetR package, **Figure 4** was generated. Several graphical modifications, that are detailed in the code, were made for the reasons of brevity and comprehension.

2.3.2 Bar plots

All bar plots in **Appendix A** were generated with the plot_bar function of Phyloseq. The x-axis was set as microhabitat, the fill as the families and the facet grid as the fields.

The plots were modified with the help of many commands from the ggplot2 package, mainly the traditional palette was replaced with a protanomaly-friendly one.

2.4 R packages used

Listed are all R packages used in the statistical analysis and their version at the time. The respective paper they were first introduced in is also referenced.

Phyloseq[19]	*v1.32*
ggplot2[20]	*v3.3.2*
PMCMR[21]	*v4.3*
PMCMRplus[22]	*v1.7*
Vegan[23]	*v2.5-6*
DESeq2[24]	*v1.28.1*
UpSetR[25]	*v1.4*
dplyr[26]	*v1.0.4*

3. Results

3.1 Spinach plants host a 'reduced-complexity' microbiota

A preliminary analysis of the dataset revealed that neither the 'Alpha diversity' Observed or Shannon measures, proxies of microbiota richness and evenness, respectively[16], were normally distributed (Shapiro-Wilk, p-value < 0.05), visualized in **Supplementary Figure 2**.

The Observed (**Figure 2A**) and Shannon (**Figure 2B**) indexes of both fields were then tested for microhabitat or field influence. The results confirmed the presence of a significant microhabitat effect (Kruskal-Wallis rank-sum test, p-value < 0.05), whereas there was no significant field effect (p-value > 0.05). This was true for both Observed and Shannon.

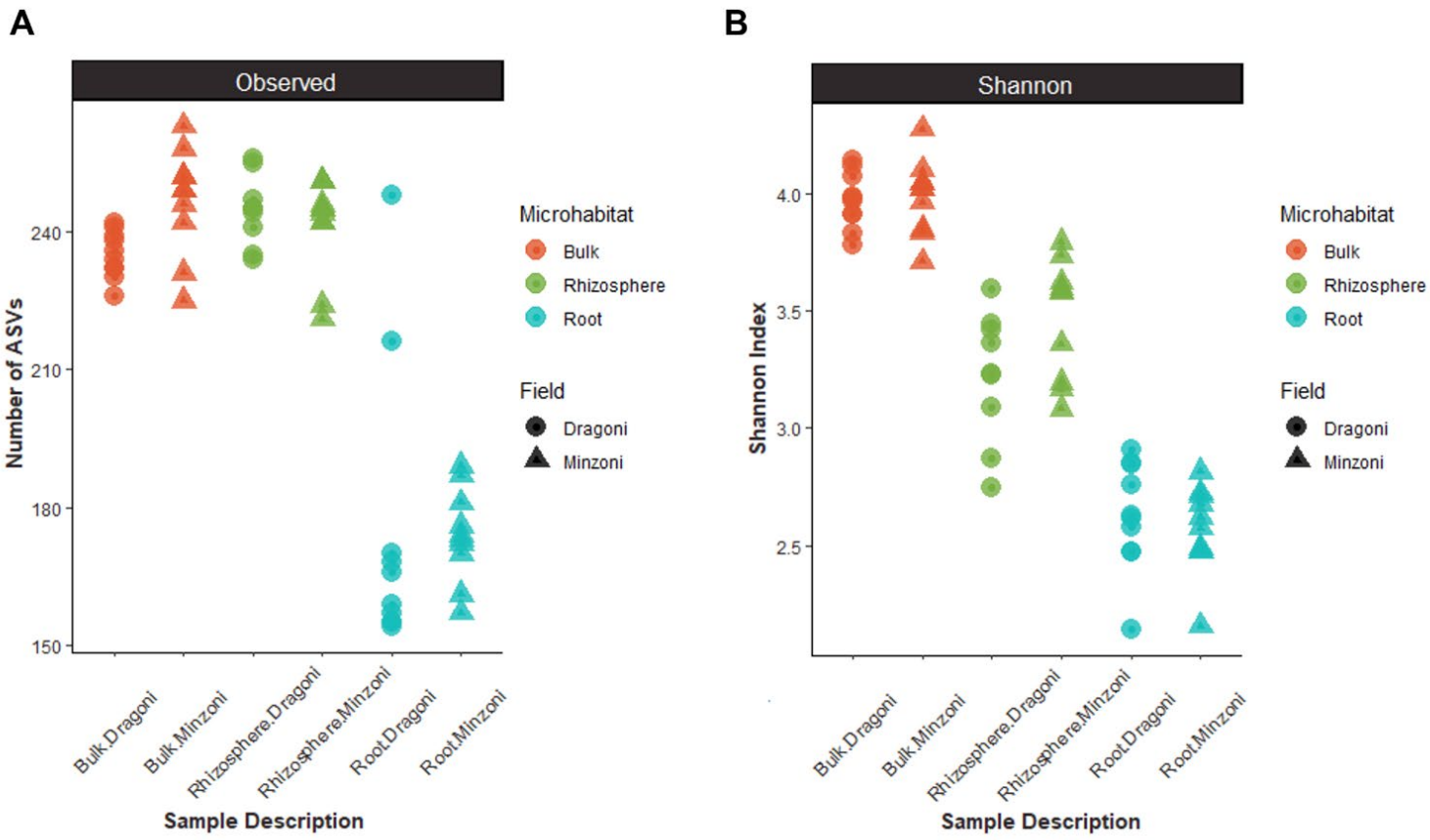


Figure 2: Alpha Diversity Measures
Individual shapes depict individual composite specimens, colour-coded according the microhabitat they were sampled form and their field of origin In A) the y-axis consists of the Number of ASVs of the specimen, in B) the y-axis consists of the Shannon index given specimens The x-axis of both A) and B) summarises the descriptions of the samples.

A PostHoc Kruskal-Dunn test[27] was used to analyse the variance pairwise. In Observed, it was found that rhizosphere and bulk microhabitats do not significantly differ (p-value > 0.05), whereas root-bulk, and root-rhizosphere did significantly differ (p-value < 0.05)In the Shannon analysis, all microhabitats did significantly differ from each other (p-value < 0.05), **Figure 2B**. This suggests that not all bacteria present in soil can equally proliferate in association with spinach.

3.2 The 'microhabitat' drives the spinach bacterial microbiota while the 'field' fine-tunes its composition

To gain further insights into the impact of the various parameters on the taxonomic composition of the spinach microbiota, I assessed the between-samples diversity, which is also known as 'Beta Diversity'[16]. First, I visualised sample compositional diversity using a canonical analysis of principal coordinates (CAP), **Figure 3**.

Congruently, a permutational analysis of variance was computed on a Bray-Curtis matrix that is sensitive to individual taxa presence and relative abundances. Interestingly, I found that the microhabitat factor, i.e., unplanted soil, rhizosphere or root was responsible for the largest amount of variance, at approximately 77.9% (adonis $R^2 = 0.7795$). Field accounted

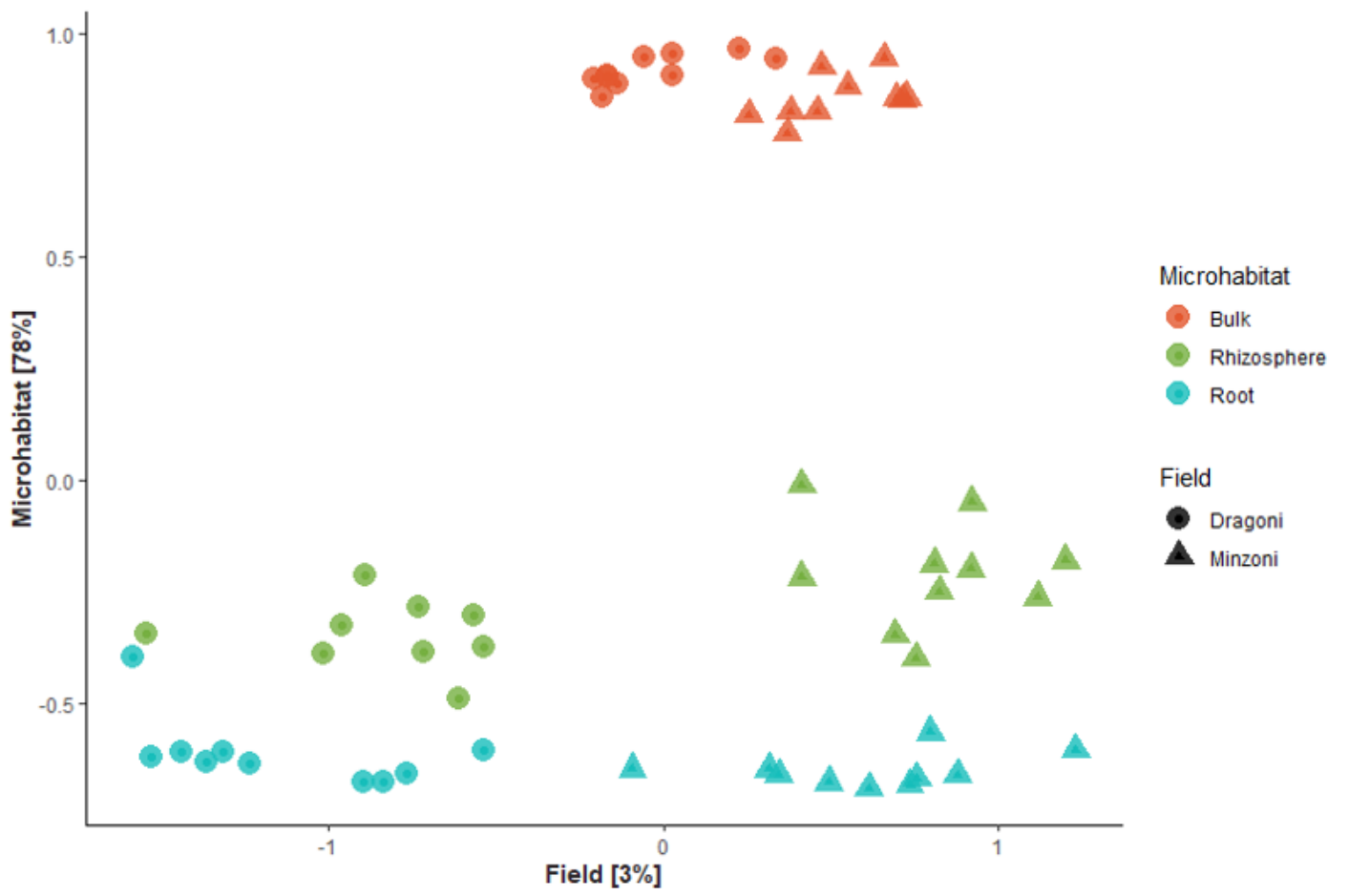


Figure 3: Beta Diversity analysis
Individual shapes depict individual composite specimens, colour-coded according the microhabitat they were sampled form ad their field of origin. In this graphical representation, the distance between two shapes is directly correlated to their microbiota diversity according to the Bray-Curtis dissimilarity matrix.

for 3.1% (adonis R^2 = 0.03109), and lastly their interaction for around 2.8% (R^2 = 0.02813) of variance in Beta diversity.

All factors had a significant effect on Beta diversity (adonis, p-value < 0.05, 5,000 permutations, **Figure 3**), while the residual unexplained variance was at 16.1% (R^2 = 0.1625).

Taken together, this analysis revealed that spinach plants host a microbiota whose taxonomic composition is different from the surrounding unplanted soil. This selective process begins in the rhizosphere and it is more marked for the communities proliferating in

roots. Interestingly, in the tested conditions, this effect is very strong and tend to “mask” the effect on different farms on the microbiota.

3.3 Identification of differentially enriched bacterial families by DESeq analysis

To gain further insights into the bacterial taxa responsible for the observed microbiota diversification at the microhabitat and field level, I identified the bacteria significantly enriched in the plant microhabitats in both farms using a DESeq analysis (**Supplementary Figure 1)**, whose output is visualized in **Figure 4**.

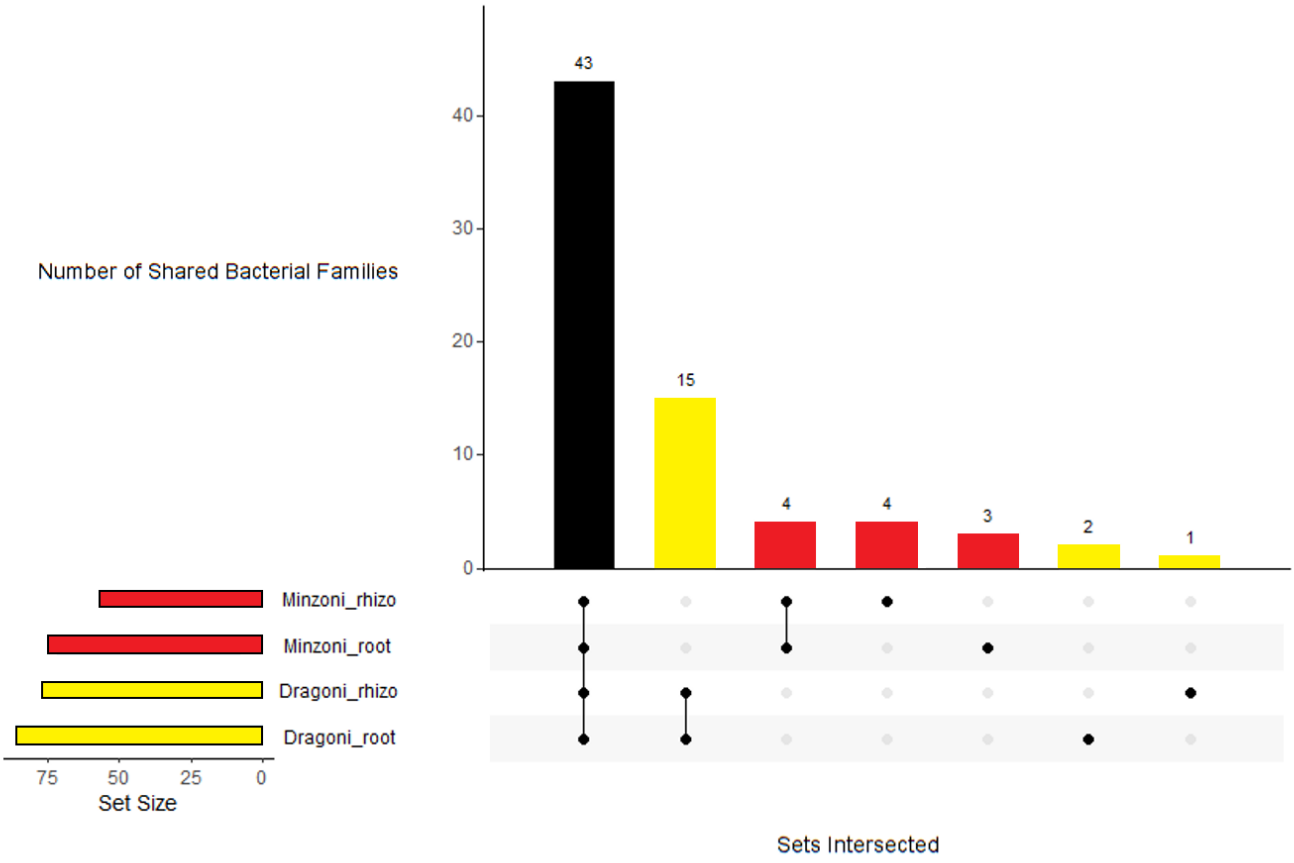


Figure 4: UpsetR plot for visualization of sets
This figure shows the amount of bacterial families enriched in a set. There are four dots, that correspond to sets in the bottom left. The intersections of that set, or the unique enrichment within that set is shown depending on the dot(s) below the bar. The black bar corresponds to the core spinach microbiome, the yellow bars to the core Dragoni microbiome, and the red bars to the core Minzoni microbiome.

Congruently with the CAP analysis, the majority of bacteria enriched in the rhizosphere and root samples (Wald test, p-value < 0.05, FDR corrected) appeared shared across fields: in total 38 individual genera belonging to 22 families were identified as representative of this category, with members of the phylum Proteobacteria (e.g., Comamonadaceae, Rhizobiaceae, Pseudomonadaceae) being the most abundant (**Figure 5**).

Interestingly, spinach plants grown in the 'Dragoni' field, which was the one characterised

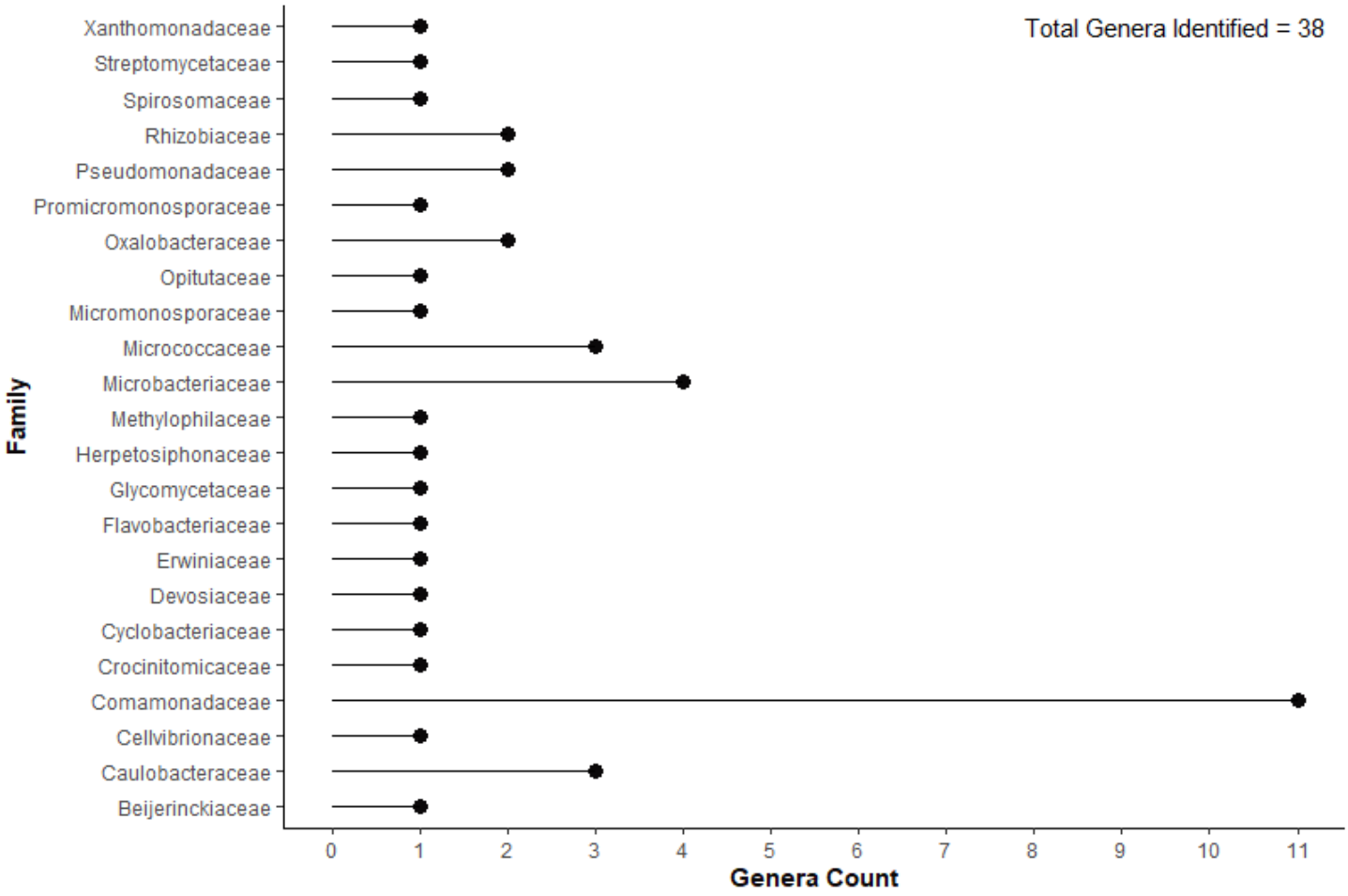


Figure 5: Bacterial families belonging to the core spinach microbiome
*The individual bacterial family is displayed on the y axis, while the number of genera corresponding to it are displayed on the x axis. This Figure summarises the taxonomic composition of the black bar in **Figure 4**.*

by the highest crop yield, were the ones hosting the more distinct microbiota. This microbiota

was still dominated by Proteobacteria but members of the family Oxalobacteraceae replaced members of the family Comamonadaceae at the top of the ranking (**Figure 6**).

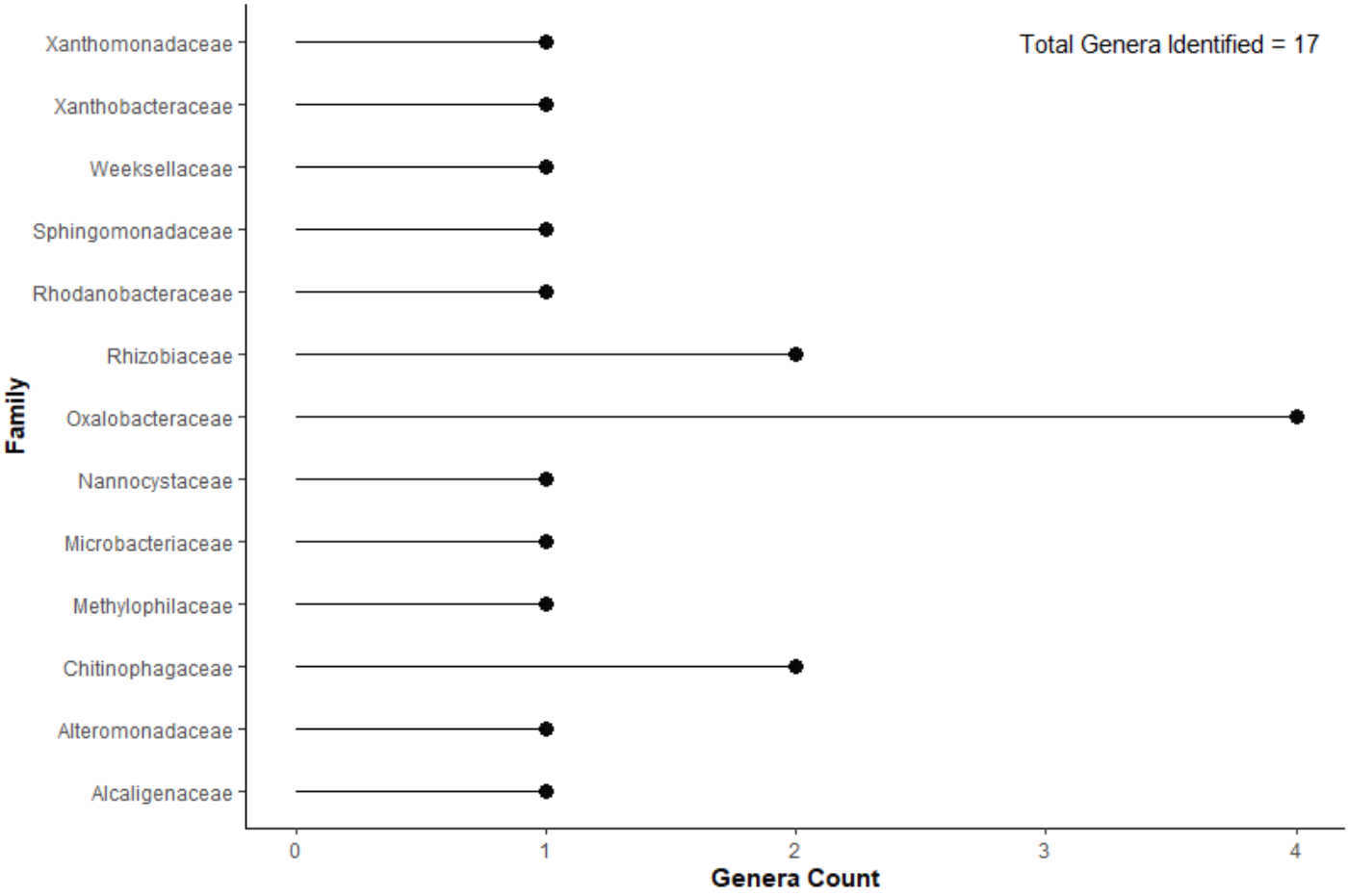


Figure 6: Bacterial families belonging to the core Dragoni microbiome
*The individual bacterial family is displayed on the y axis, while the number of genera corresponding to it are displayed on the x axis. This Figure summarises the taxonomic composition of the yellow bars in **Figure 4**.*

Finally, I determined the taxonomic composition of the 'Minzoni' field. This appeared more congruent with the core spinach microbiome as members of the family Comamonadaceae were still dominant in the community (**Figure 7**).

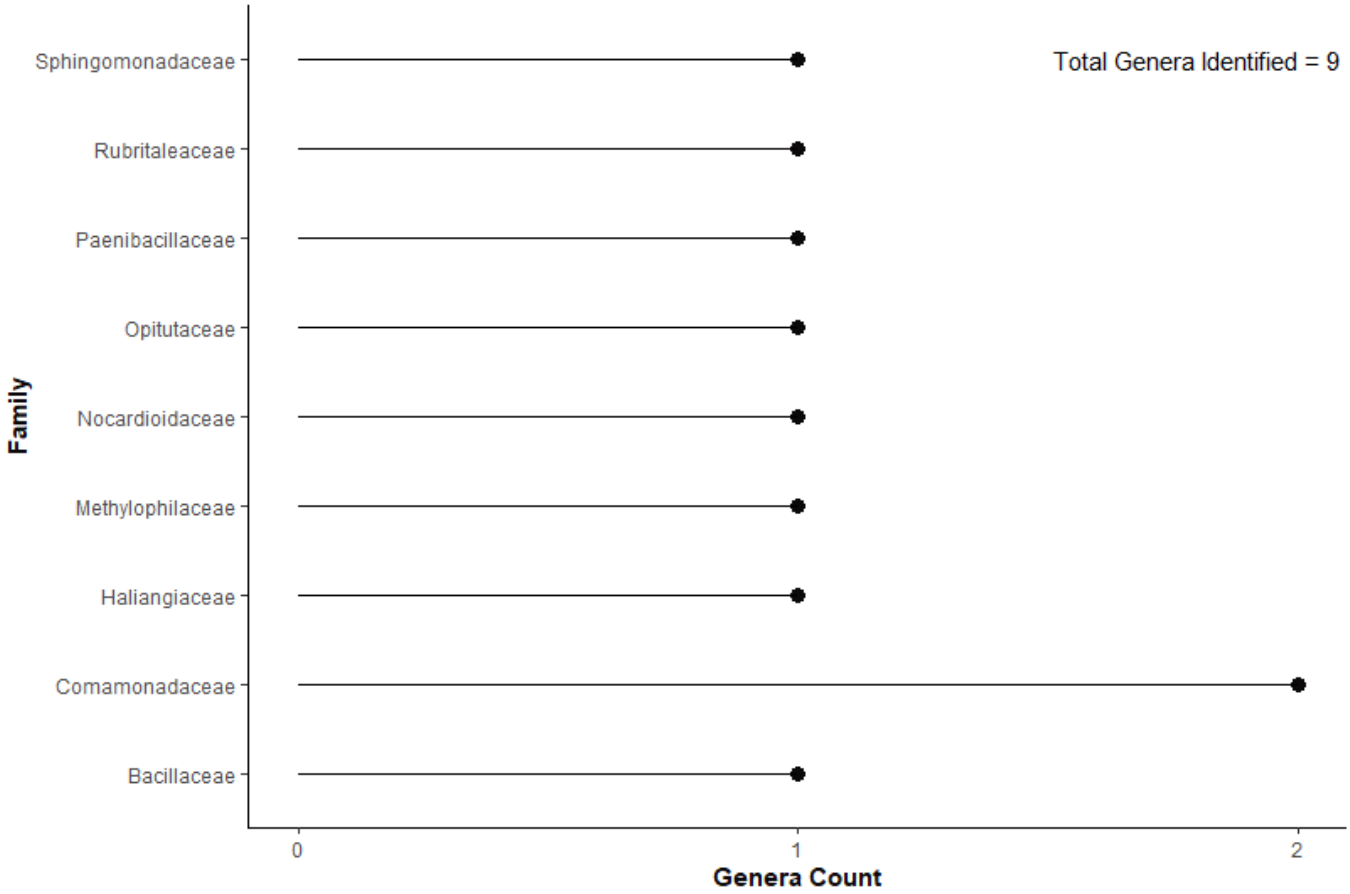


Figure 7: Bacterial families belonging to the core Minzoni microbiome
The individual bacterial family is displayed on the y axis, while the number of genera corresponding to it are displayed on the x axis. This Figure summarises the taxonomic composition of the red bars in ***Figure 4****.*

3.4 Genera corresponding to the enriched bacterial families

As the bacterial families of the taxa were identified, so were the genera belonging to it. Note that some reads were assigned NA at the Genus level, which is the reason for the apparent discrepancy between the total genera identified (**Figures 5, 6** and **7**) and the number of differentially enriched bacteria (**Figure 4**).

3.4.1 Genera in the core spinach microbiome

The genera belonging to the core spinach microbiome are listed in **Table 1**.

Family	Genus
Beijerinckiaceae	Bosea
Caulobacteraceae	Caulobacter
Caulobacteraceae	Asticcacaulis
Caulobacteraceae	Phenylobacterium
Cellvibrionaceae	Cellvibrio
Comamonadaceae	Limnohabitans
Comamonadaceae	Rhizobacter
Comamonadaceae	Polaromonas
Comamonadaceae	Paucibacter
Comamonadaceae	Pseudorhodoferax
Comamonadaceae	Rhodoferax
Comamonadaceae	Azohydromonas
Comamonadaceae	Leptothrix
Comamonadaceae	Hydrogenophaga
Comamonadaceae	Variovorax
Crocinitomicaceae	Fluviicola
Devosiaceae	Devosia
Erwiniaceae	Pantoea
Flavobacteriaceae	Flavobacterium
Glycomycetaceae	Glycomyces
Herpetosiphonaceae	Herpetosiphon
Methylophilaceae	Methylotenera
Microbacteriaceae	Agromyces
Microbacteriaceae	Galbitalea
Microbacteriaceae	Microbacterium
Micrococcaceae	Pseudarthrobacter
Micrococcaceae	Paenarthrobacter
Micromonosporaceae	Actinoplanes
Opitutaceae	Lacunisphaera
Oxalobacteraceae	Massilia
Oxalobacteraceae	Duganella
Promicromonosporaceae	Promicromonospora
Pseudomonadaceae	Pseudomonas
Rhizobiaceae	Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium
Rhizobiaceae	Shinella
Spirosomaceae	Dyadobacter
Streptomycetaceae	Streptomyces
Xanthomonadaceae	Stenotrophomonas

Table 1: core spinach microbiome genera

The bacterial families enriched in the core spinach microbiome are listed along with their corresponding genus.

3.4.2 Genera in the core Dragoni microbiome

The genera belonging to the core Dragoni microbiome are listed in **Table 2**.

Family	Genus
Alcaligenaceae	Achromobacter
Alteromonadaceae	Rheinheimera
Chitinophagaceae	Cnuella
Chitinophagaceae	Aurantisolimonas
Methylophilaceae	MM2
Microbacteriaceae	Frondihabitans
Nannocystaceae	Nannocystis
Oxalobacteraceae	Oxalobacter
Oxalobacteraceae	Oxalicibacterium
Oxalobacteraceae	Herminiimonas
Oxalobacteraceae	Herbaspirillum
Rhizobiaceae	Mesorhizobium
Rhizobiaceae	Phyllobacterium
Rhodanobacteraceae	Ahniella
Sphingomonadaceae	Sphingopyxis
Weeksellaceae	Chryseobacterium
Xanthomonadaceae	Luteimonas

Table 2: core Dragoni microbiome genera
The bacterial families enriched in the core Dragoni microbiome are listed along with their corresponding genus.

3.4.3 Genera in the core Minzoni microbiome

The genera belonging to the core Minzoni microbiome are listed in **Table 3**.

Family	Genus
Comamonadaceae	Ramlibacter
Comamonadaceae	Acidovorax
Haliangiaceae	Haliangium
Methylophilaceae	Methylophilus
Nocardioidaceae	Kribbella
Opitutaceae	IMCC26134
Paenibacillaceae	Paenibacillus
Rubritaleaceae	Luteolibacter
Sphingomonadaceae	Sphingobium

Table 3: core Minzoni microbiome genera
The bacterial families enriched in the core Minzoni microbiome are listed along with their corresponding genus.

4. Discussion

In this section, I will discuss the mechanisms driving the selection of the spinach microbiome, infer its functions and illustrate potential avenues for applications in agriculture.

4.1 What may drive the selection of the spinach microbiome under field conditions?

The core spinach microbiome exhibits the characteristics of a 'gated community'. In which the closer we approach the roots of a plant, the less diversity there is[28]. This is a "hallmark" of the plant microbiota which has been identified in many other plant species[10]. As the soil pH of the two farms was comparable (~pH 8, Bulgarelli D., personal communication), and soil pH is a key determinant of bacterial microbiota composition[29]. It is plausible that other farm factors drive the observed spinach microbiota diversification. Although the identification of the mechanisms that a spinach plant recruits bacteria by was outside the scope of this dissertation, I speculate that genotypic differences between the tested cultivars, for example, those at the level of rhizodeposition, RSA or immune responses, contribute at least partially to the observed 'Field effect'. As genomic resources for Spinach are becoming increasingly available[30], it will be possible to test, in future experimentations, whether some of these genotypic differences contribute to microbiome assembly.

For example, whether the introduction of the spinach plant to the soil results in the generation of an environment where microbes beneficial to the plant can outcompete other non-beneficial or pathogenic microbes.

As the project will contemplate additional sequencing datasets, future investigations may profit from advanced analytical techniques, such as machine learning[31], to identify and

quantify individual factors (e.g., soil characteristics, seasonal variation, plant genotype, etc.) putatively associated with microbiome composition

4.2 Members of the spinach microbiome have the potential to be plant-growth-promoting rhizobacteria (PGPRs)

The core spinach microbiome seems to be predominantly comprised of three bacterial families (Supplementary **Figure 3B**): Oxalobacteraceae, Pseudomonadaceae, and Streptomycetaceae.

Oxalobacteraceae has been known to contain members with nitrogen-fixing or antibacterial properties[32]. Out of the two genera identified, bacteria belonging to the genus *Massilia* have been shown to reduce the presence of a broad spectrum of gram-positive bacteria and fungi in the soil[32].

Previous research has characterized the plant growth-promoting nature of the family Pseudomonadaceae, especially of the *Pseudomonas* genus[33,34]. Species belonging to the *Pseudomonas* genus have been known to secrete insecticidal toxins, reduce the growth of harmful pathogens in the soil, and some of the antibiotics secreted by it have positively impacted attempts to manage nematode infestations[34].

Reads belonging to the important *Streptomyces* genus of Streptomycetaceae were also identified. *Streptomyces* species have shown potential as PGPRs in the past due to their ability to produce plant growth regulators and their affinity for the production of novel antibiotics[35]. Previous attempts have been successful in generating *Streptomyces* species with antipathogenic properties and then using them to positively impact the growth of coniferous plants[36].

As all of these have shown PGPR properties and are found in high abundances within the core spinach microbiome (**Supplementary Figure 4**), they may be of interest for future research trying to develop universal spinach inoculant PGPRs.

4.3 Plants in Dragoni Field differentially recruit a bacterial family shown to be a sign of healthy soil

It has been proposed that Chitinophagaceae can be used as a microbial predictor of soil health[37]. This is due to them being negatively correlated to the presence of aluminium in the soil[37]. Aluminium is a toxin known to significantly limit crop productivity when in acidic environments[38].

More importantly than Chitinophagaceae's negative correlation with Aluminium[37] and significant enrichment within the core Dragoni microbiome (**Supplementary Figure 3A**) is the fact that it is not found at a significant amount in the core Minzoni microbiome (**Supplementary Figure 3B**). This suggests that the Minzoni field may be suffering from aluminium stress that is not present in the Dragoni field.

4.4 The Minzoni field may have the presence of plant pathogens

The *Acidivorax* genera of the bacterial family Comamonadaceae are known to exhibit plant-pathogenic properties[39]. Reads belonging to the *Acidivorax* and *Ramlibacter* genera were found to be significantly enriched in the core Minzoni microbiome as compared to core Dragoni microbiome (**Supplementary Figures 3A and 3B**).

This suggests that there is a presence of putative pathogenic *Acidivorax* species in the Minzoni field that infect the Spinach plant when it is introduced. This may be due to the environmental conditions of the Minzoni field being more favourable for *Acidivorax*

pathogenesis or the cultivar planted in the Minzoni field being more susceptible to infection from this genus.

4.5 The limitations of this study

It is important to note when discussing the properties of bacterial family and genera, that there are always some species that do not conform to or have present within themselves the property in question. The inability to identify the species the read belongs to is unfortunately the limitation of the sequencing technology available to us[40]. Likewise, the lack of either the soil metagenome[41] or whole-genome sequencing information[42] for representative strains of the bacteria identified in the sequencing survey prevents me from firmly concluding on the metabolic capacities of the spinach microbiome. A way to address this is to use other sequencing approaches, such as metagenomics, which can reveal insights into the functionality of the individual reads[43]. Thus, allowing us to narrow the species even further.

Additionally, I am unable to confirm whether the observed microbiota diversification is due to the microbes enabling the plant to perform better, or due to the possibility that plants that grow better then tend to host different microbiotas. This can be resolved with the implementation of an inoculation assay[44]. For instance, where the significantly enriched microbes identified in the core Dragoni microbiome are prepared as an inoculant and applied to Minzoni recipient plants. The impact of the inoculation is then monitored at the level of the plant performance to resolve this issue.

Regardless of the outcome of these future experiments, the sequencing information retains a diagnostic value for spinach production (further elaborated upon in **Section 5**).

However, before implementing the inoculation experiment discussed above, it will need to be identified whether Spinach plants contain seed-transmitted endophytes[45]. If any seed-transmitted endophytes are discovered, they will need to be identified and compared against the bacteria contained within the soil. This essential for determining whether the target of the inoculant will be the seeds or the soil.

Results from this dissertation are also limited to the Emilia Romagna region in Italy and one sampling season. Additionally, the cultivars used for the fields were not the same. The genotype of the host plant has been shown affect the microbiota hosted[12] and experiments where confounding factors can be firmly separated (e.g., different genotypes grown in the same field or the same genotype grown in different fields) are required. Of note, these approaches were integrated into the subsequent sampling campaigns of the project.

5. Conclusion

In this work, I have identified that as the fields differ in crop productivity, they also do so in the content of their microbiome. Additionally, I have identified bacteria significantly differing between the two fields along with the bacterial Family and (in some cases) the Genus they belong to, as well as the abundances of said families. Differentially enriched bacterial families of interest have also been discussed as having potentially PGPR, soil predicting, or pathogenic qualities.

It is critical that any seed-borne endophytes are identified before the results detailed in this study can be utilized. It is my suggestion that future research focuses on this before building upon this paper's findings. Likewise, a better appreciation of the overall genotype effect as well as the contribution of other members of the microbiome, e.g., fungi, will be required to infer first principles.

Yet, as an immediate outcome, the differentially recruited bacteria can be tested as diagnostic markers to predict the yield potential of certain fields, similar to what has been proposed for disease suppressive soils [46]. Furthermore, after confirmation of this research with inoculation assays, the experimental approach used here can be adapted to target the microbiome of other crops phylogenetically related to spinach, such as memebers of the Amaranthaceae family including *Beta vulgaris* (Sugar beet) or *Chenopodium quinoa* (Quinoa).

By 2050 the world's population is expected to exponentially grow from 7.7 billion to 9.7 billion[47]. As it stands, our current agricultural methods are incapable of sustaining such an increase. This research is just a part of a larger body of work being carried out by the Circles EU project to address global food supply and other sustainable developmental goals set by the UN[48].

6. Acknowledgements

I thank Dr Davide Bulgarelli for going above and beyond his requirements as an Honours Supervisor to support me during this dissertation. Additionally, I also acknowledge the help and feedback of the members of the Davide Bulgarelli lab (@Team_DB) at the James Hutton Institute, particularly Dr Senga Robertson-Albertyn for her feedback and out-of-hours support. Finally, I thank all my friends and family for encouraging and pushing me to complete this dissertation during a challenging time.

7. Funding

The sequencing dataset analysed in this dissertation was funded as part of the H2020 project ‘CIRCLES’ (Grant Agreement 818920).

8. References

1. Hunter MC, Smith RG, Schipanski ME, Atwood LW, Mortensen DA. Agriculture in 2050: Recalibrating targets for sustainable intensification. Bioscience. 2017;67(4):386–91.

2. Birkby. Vertical Farming | AProgram. ATTRA Sustain Agric [Internet]. 2016;1–12. Available from: https://attra.ncat.org/product/Vertical-Farming/

3. William F G. HYDROPONICS—CROP PRODUCTION IN LIQUID CULTURE MEDIA. Science (80-). 1937;85(2198):177–8.

4. Hickey LT, N. Hafeez A, Robinson H, Jackson SA, Leal-Bertioli SCM, Tester M, Gao C, Godwin ID, Hayes BJ, Wulff BBH. Breeding crops to feed 10 billion. Nat Biotechnol [Internet]. 2019;37(7):744–54. Available from: http://dx.doi.org/10.1038/s41587-019-0152-9

5. Finkel OM, Castrillo G, Herrera Paredes S, Salas González I, Dangl JL. Understanding and exploiting plant beneficial microbes. Curr Opin Plant Biol. 2017;38:155–63.

6. Compant S, Samad A, Faist H, Sessitsch A. A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. J Adv Res. 2019;19:29–37.

7. Trivedi P, Schenk PM, Wallenstein MD, Singh BK. Tiny Microbes, Big Yields: enhancing food crop production with biological solutions. Microb Biotechnol. 2017;10(5):999–1003.

8. Bhattacharyya PN, Jha DK. Plant growth-promoting rhizobacteria (PGPR): Emergence in agriculture. World J Microbiol Biotechnol. 2012;28(4):1327–50.

9. Kumar A, Verma JP. The Role of Microbes to Improve Crop Productivity and Soil Health. In: Ecological Wisdom Inspired Restoration Engineering. Springer, Singapore; 2018. p. 249–65.

10. Bulgarelli D, Schlaeppi K, Spaepen S, Van Themaat EVL, Schulze-Lefert P. Structure and functions of the bacterial microbiota of plants. Annu Rev Plant Biol. 2013;64:807–38.

11. Alegria Terrazas R, Giles C, Paterson E, Robertson-Albertyn S, Cesco S, Mimmo T, Pii Y, Bulgarelli D. Plant-microbiota interactions as a driver of the mineral Turnover in the Rhizosphere [Internet]. Vol. 95, Advances in Applied Microbiology. Elsevier Ltd; 2016. 1–67 p. Available from: http://dx.doi.org/10.1016/bs.aambs.2016.03.001

12. Escudero-Martinez C, Bulgarelli D. Tracing the evolutionary routes of plant–microbiota interactions. Curr Opin Microbiol. 2019;49:34–40.

13. Schlaeppi K, Bulgarelli D. The plant microbiome at work. Mol Plant-Microbe Interact. 2015;28(3):212–7.

14. Nair P. Woese and Fox: Life, rearranged. Proc Natl Acad Sci U S A. 2012;109(4):1019–21.

15. Caradonia F, Ronga D, Catellani M, Giaretta Azevedo CV, Terrazas RA, Robertson-Albertyn S, Francia E, Bulgarelli D. Nitrogen fertilizers shape the composition and predicted functions of the microbiota of field-grown tomato plants. Phytobiomes J. 2019;3(4):315–25.

16. Whittaker RH. Evolution and Measurement of Species Diversity. Taxon. 1972;21(2):213–51.

17. Robertson-Albertyn S, Terrazas RA, Balbirnie K, Blank M, Janiak A, Szarejko I,

Chmielewska B, Karcz J, Morris J, Hedley PE, George TS, Bulgarelli D. Root hair mutations displace the barley rhizosphere microbiota. Front Plant Sci. 2017;8(June):1–15.

18. Alegria Terrazas R, Balbirnie-Cumming K, Morris J, Hedley PE, Russell J, Paterson E, Baggs EM, Fridman E, Bulgarelli D. A footprint of plant eco-geographic adaptation on the composition of the barley rhizosphere bacterial microbiota. Sci Rep [Internet]. 2020;10(1):1–13. Available from: https://doi.org/10.1038/s41598-020-69672-x

19. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8(4).

20. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K, Yutani H, Dunnington D. ggplot2: Elegant Graphics for Data AnalysisNo Title [Internet]. Springer-Verlag New York; 2016. Available from: https://ggplot2.tidyverse.org

21. Pohlert T. The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR) [Internet]. 2014. Available from: https://cran.r-project.org/package=PMCMR

22. Pohlert T. PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums [Internet]. 2020. Available from: https://cran.r-project.org/web/packages/PMCMRplus/PMCMRplus.pdf

23. Oksanen AJ, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, Minchin PR, Hara RBO, Simpson GL, Solymos P, Stevens MHH, Szoecs E. Package ‘ vegan .’ 2020.

24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):1–21.

25. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of

Intersecting Sets Europe PMC Funders Group. IEEE Trans Vis Comput Graph [Internet]. 2014;20(12):1983–92. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4720993/pdf/emss-66749.pdf

26. Wickham H, François R, Henry L, Müller K. Package ‘ dplyr .’ 2021.

27. Dunn OJ. Multiple Comparisons Using Rank Sums. Technometrics. 1964;6(3):241–52.

28. Naylor D, Fansler S, Brislawn C, Nelson WC, Hofmockel KS, Jansson JK, McClure R. Deconstructing the Soil Microbiome into Reduced-Complexity Functional Modules. MBio. 2020;11(4):1–19.

29. Lauber CL, Hamady M, Knight R, Fierer N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol. 2009;75(15):5111–20.

30. Xu C, Jiao C, Sun H, Cai X, Wang X, Ge C, Zheng Y, Liu W, Sun X, Xu Y, Deng J, Zhang Z, Huang S, Dai S, Mou B, Wang Q, Fei Z, Wang Q. Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. Nat Commun [Internet]. 2017;8(May):1–10. Available from: http://dx.doi.org/10.1038/ncomms15275

31. Namkung J. Machine learning methods for microbiome studies. J Microbiol. 2020;58(3):206–16.

32. Baldani JI, Rouws L, Cruz LM, Olivares FL, Schmid M, Hartmann A. 35 The Family Oxalobacteraceae. In: The Prokaryotes – Alphaproteobacteria and Betaproteobacteria. Springer-Verlag Berlin Heidelberg; 2014. p. 919–74.

33. Palleroni NJ. 58 Introduction to the Family Pseudomonadaceae. In: Starr MP, Stolp H, Trüper HG, Balows A, Schlegel HG, editors. The Prokaryotes: A Handbook on

Habitats, Isolation, and Identification of Bacteria. Springer-Verlag Berlin Heidelberg; 1981. p. 655–65.

34. Sankari Meena K, Annamalai M, Prabhukarthikeyan SR, Keerthana U, Yadav MK, Rath PC, Jena M, Prajna P. Agriculture Application of Pseudomonas: A View on the Relative Antagonistic Potential Against Pests and Diseases. Plant Growth Promot Rhizobacteria Agric Sustain. 2019;77–93.

35. Olanrewaju OS, Babalola OO. Streptomyces: implications and interactions in plant growth promotion. Appl Microbiol Biotechnol. 2019;103(3):1179–88.

36. de Vasconcellos RLF, Cardoso EJBN. Rhizospheric streptomycetes as potential biocontrol agents of Fusarium and Armillaria pine rot and as PGPR for Pinus taeda. BioControl. 2009;54(6):807–16.

37. Hermans SM, Buckley HL, Case BS, Curran-cournane F, Taylor M, Lear G. Bacteria as Emerging Indicators of Soil Condition. Appl Environ Microbiol. 2017;83(1):1–13.

38. Mota D, Faria F, Gomes EA, Paiva IEME, Seldin L. Bacterial and Fungal Communities in Bulk Soil and Rhizospheres of Aluminum-Tolerant and Aluminum-Sensitive Maize (Zea mays L.) Lines Cultivated in Unlimed and Limed Cerrado Soil. J Microbiol Biotechnol. 2008;18(5):805–14.

39. Willems A. 29 The Family Comamonadaceae. In: Rosenberg E, editor. The Prokaryotes: Alphaproteobacteria and Betaproteobacteria. Springer-Verlag Berlin Heidelberg; 2014. p. 777–851.

40. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. Nat Methods. 2010;8(1):61–5.

41. Nesme J, Achouak W, Agathos SN, Bailey M, Baldrian P, Brunel D, Frostegård Å,

Heulin T, Jansson JK, Jurkevitch E, Kruus KL, Kowalchuk GA, Lagares A, Lappin-Scott HM, Lemanceau P, Le Paslier D, Mandic-Mulec I, Murrell JC, Myrold DD, Nalin R, Nannipieri P, Neufeld JD, O'Gara F, Parnell JJ, Pühler A, Pylro V, Ramos JL, Roesch LFW, Schloter M, Schleper C, Sczyrba A, Sessitsch A, Sjöling S, Sørensen J, Sørensen SJ, Tebbe CC, Topp E, Tsiamis G, Van Elsas JD, Van Keulen G, Widmer F, Wagner M, Zhang T, Zhang X, Zhao L, Zhu YG, Vogel TM, Simonet P. Back to the future of soil metagenomics. Front Microbiol. 2016;7(FEB):1–5.

42. Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, Rangel Alvarez B, Lundberg DS, Lu TY, Lebeis S, Jin Z, McDonald M, Klein AP, Feltcher ME, Rio TG, Grant SR, Doty SL, Ley RE, Zhao B, Venturi V, Pelletier DA, Vorholt JA, Tringe SG, Woyke T, Dangl JL. Genomic features of bacterial adaptation to plants. Nat Genet [Internet]. 2018;50(1):138–50. Available from: http://dx.doi.org/10.1038/s41588-017-0012-9

43. Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman JA, Hugenholtz P, Van Der Lelie D, Meyer F, Stevens R, Bailey MJ, Gordon JI, Kowalchuk GA, Gilbert JA. Unlocking the potential of metagenomics through replicated experimental design. Nat Biotechnol. 2012;30(6):513–20.

44. Vorholt JA, Vogel C, Carlström CI, Müller DB. Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. Cell Host Microbe. 2017;22(2):142–55.

45. Shahzad R, Khan AL, Bilal S, Asaf S, Lee IJ. What is there in seeds? Vertically transmitted endophytic resources for sustainable improvement in plant growth. Front Plant Sci. 2018;9:1–10.

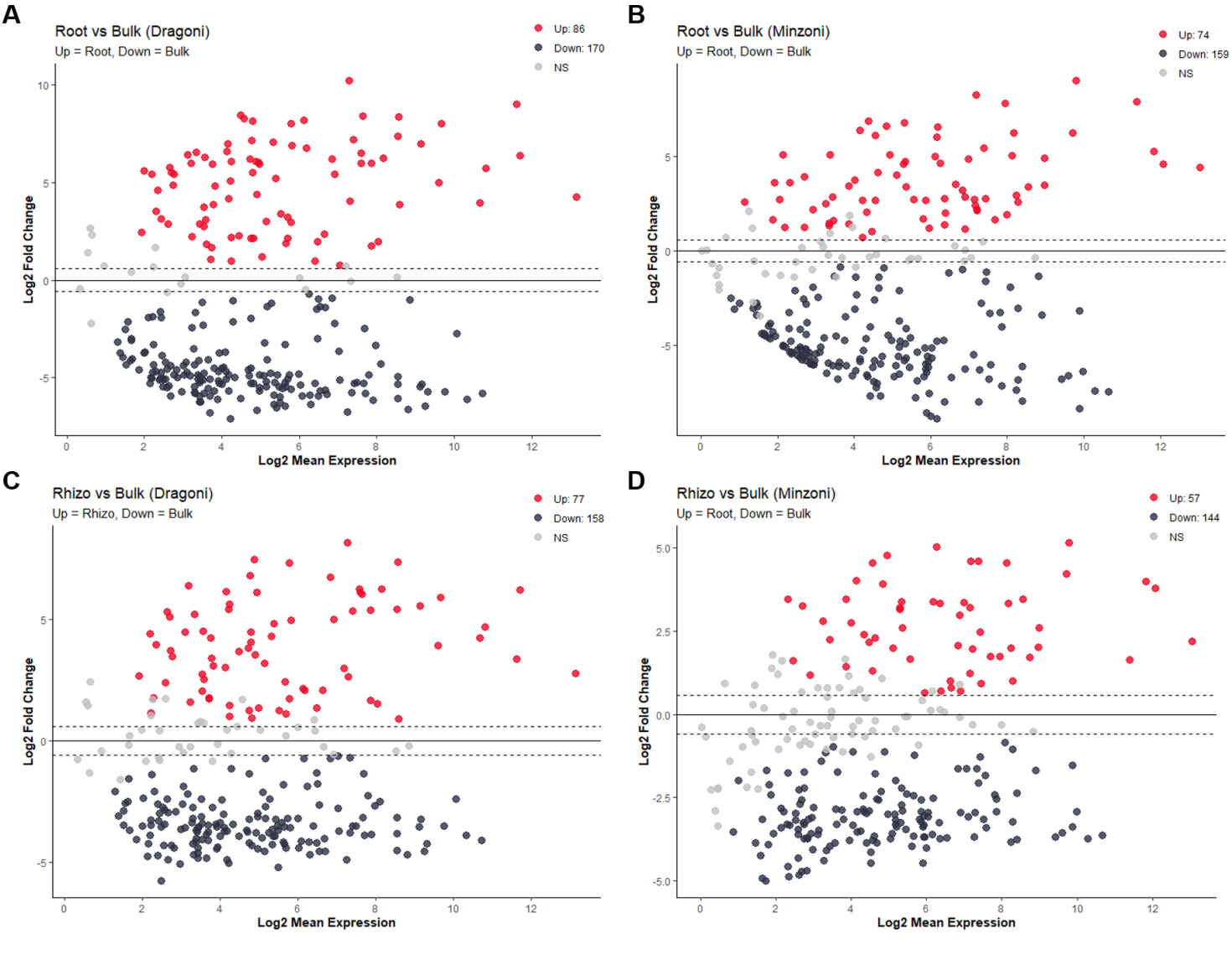
46. Yuan J, Wen T, Zhang H, Zhao M, Penton CR, Thomashow LS, Shen Q. Predicting

disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt. ISME J [Internet]. 2020;14(12):2936–50. Available from: http://dx.doi.org/10.1038/s41396-020-0720-5

47. World Population Prospects 2019, Population Data [Internet]. United Nations Population Division; 2019. p. Total Population-Both Sexes, Estimates tab. Available from: https://population.un.org/wpp/Download/Standard/Population/

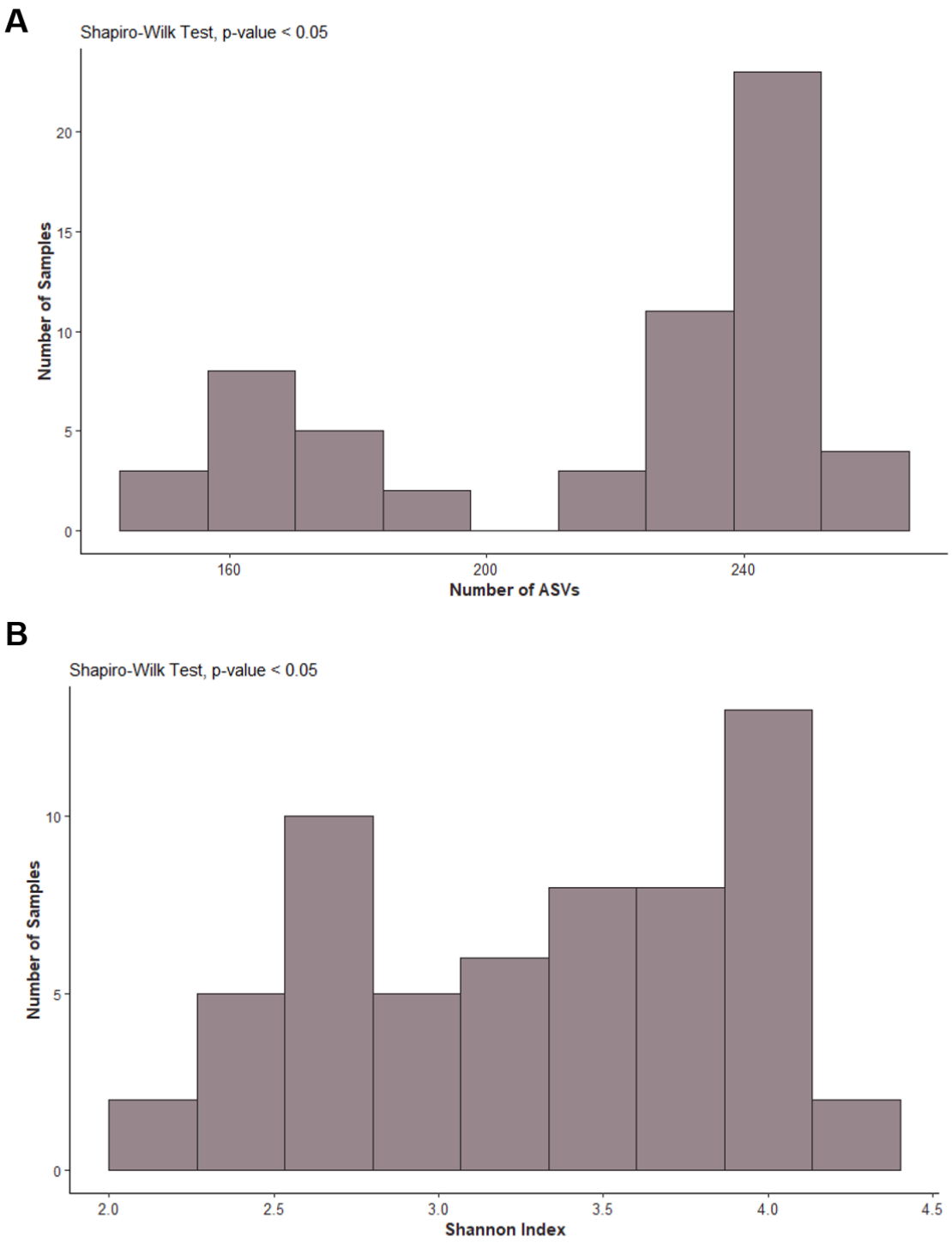
48. D'Hondt K, Kostic T, McDowell R, Eudes F, Singh BK, Sarkar S, Markakis M, Schelkle B, Maguin E, Sessitsch A. Microbiome innovations for a sustainable future. Nat Microbiol [Internet]. 2021;6(2):138–42. Available from: http://dx.doi.org/10.1038/s41564-020-00857-w

Appendix 1: Supplementary Figures



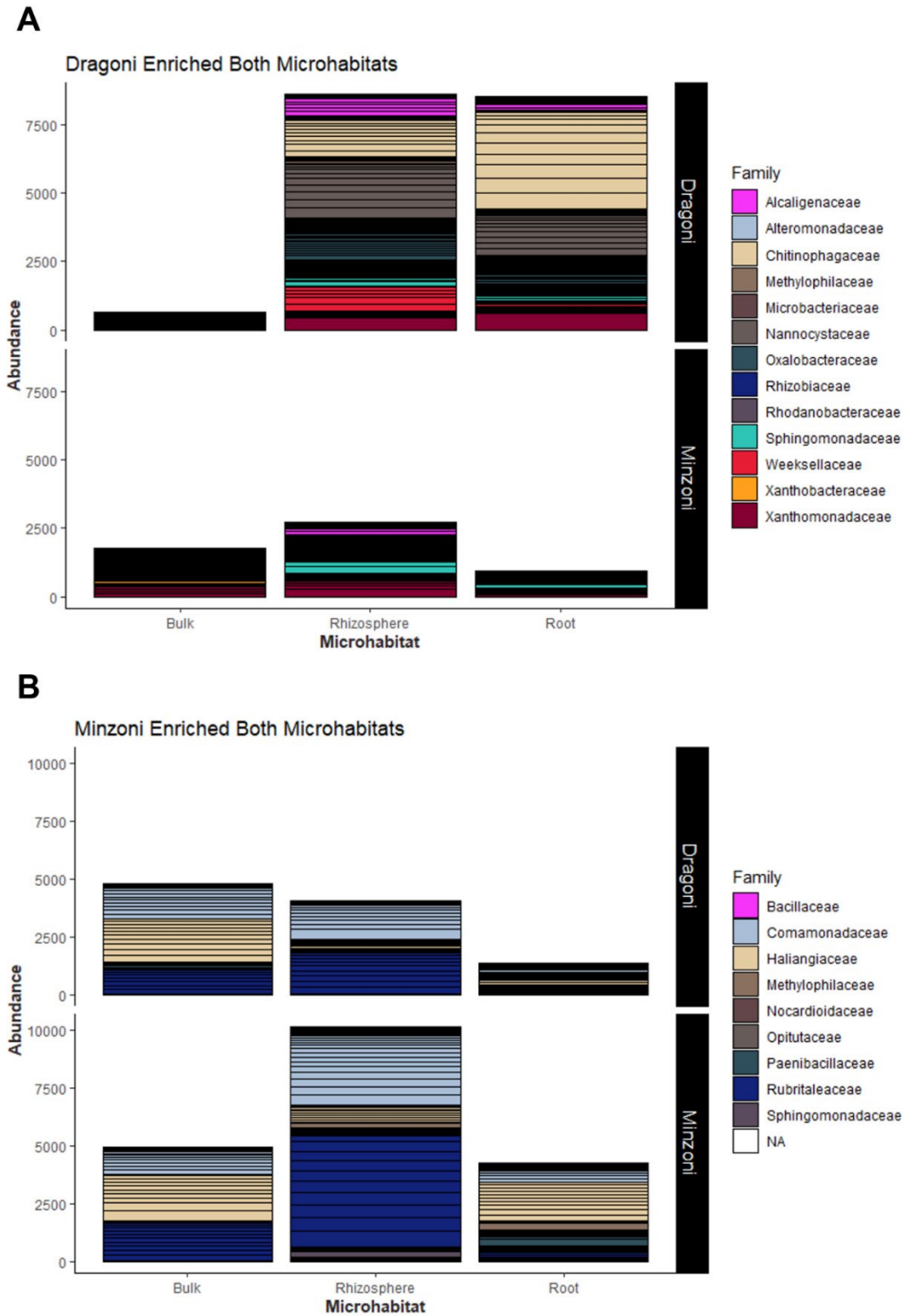
Supplementary Figure 1: MA-plots

Each dot in these MA-plots visualise bacteria differential enriched between A) Dragoni Root and Bulk, B) Minzoni Root and Bulk, C) Dragoni Rhizo and Bulk, D) Minzoni Rhizo and Bulk. The x-axis denotes bacterial abundances in sequencing reads and the y-axis the log fold change between the indicated term of comparison. Coloured dots depict significantly differentially recruited bacteria (Wald test, P value < 0.05, FDR corrected) according to the legend found in the top right corner of each graph.



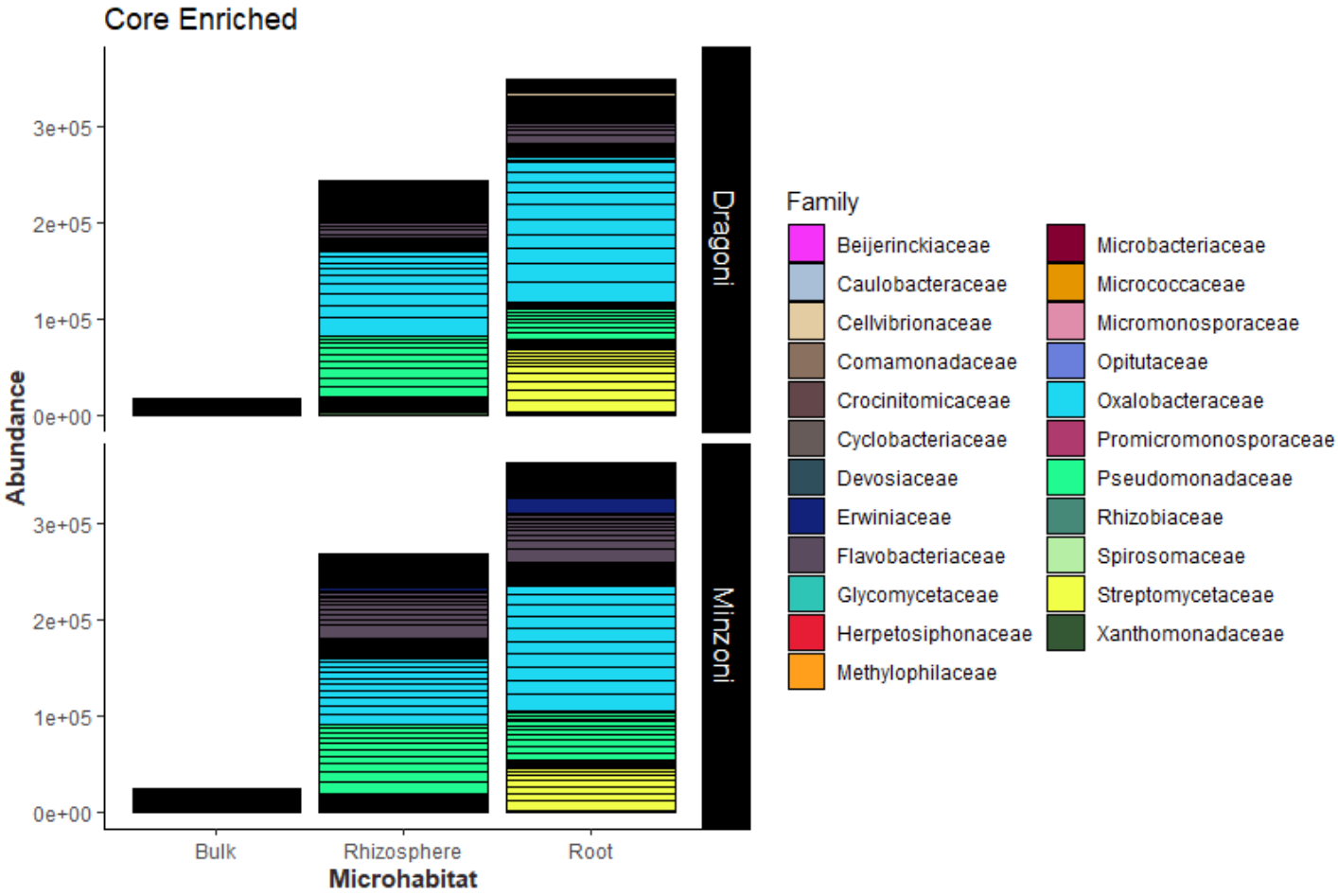
Supplementary Figure 2: Distribution of reads

The x-axis depicts the number of samples that correspond to the y-axis which is A) the Number of ASVs, and B) a value on the Shannon index.



Supplementary Figure 3: Bacteria enriched in A) the core Dragoni microbiome, and Bacteria enriched B) in the core Minzoni microbiome

On the x-axis are the microhabitats and on the y-axis are the bacterial abundance. Multiple layers of same color depict individual biological replicates. The plots are additionally split by Field to allow for comparison.



Supplementary Figure 4: Bacteria enriched in the core spinach microbiome

On the x-axis are the microhabitats and on the y-axis are the bacterial abundance. Multiple layers of same color depict individual biological replicates. This plot is additionally split by Field to allow for comparison.

Appendix 2: Reflective Account of Honors Research

(A copy of your reflective account MUST be included with your thesis)

Section 1

Student Name: *Muhammad Amanullah Jilani*

Matriculation Number: *180017551*

Email address: *mjilani@dundee.ac.uk*

Degree Specialism (if known): *BSc (Hons) Molecular Genetics*

Section 2

Principle Investigator: *Dr Davide Bulgarelli*

Laboratory Location (if applicable): *N/A*

Day-to-Day supervisor (if applicable): *Dr Davide Bulgarelli*

Section 3

1. *With respect to the Methodology/Technical Summary that you submitted in your Grant Proposal, explain how your earlier expectations been altered by your project experience?* ***[400 Words]***

Changes in understanding of research

I have gone from someone with absolutely no R-studio/coding experience to someone who is, very comfortably, able to carry out the generation of ecological indexes and compute statistical analysis on 16S rRNA Illumina sequencing datasets. This is in large part due to the guidance I received from Dr Bulgarelli and from my own experience gained by doing another project that is very similar to this.

Through diving into the research papers where the packages I used (**Section 2.4**) first appeared, I have developed an above-average understanding of the intricacies surrounding the commands of the packages I used, this will help me in addressing other, or additional, analysis in R studio.

I have also become much more knowledgeable in the specificities of the two main ecological indexes and DESeq analysis.

A notable improvement has been in learning how to generate visually effective and statistically relevant graphs and figures. My core focus has been on ensuring the reproducibility of my analysis and making this research easier to access. These two points were not at all considered when I first began this dissertation and were instead repeatedly reinforced by inevitable exposure to the good practice of others carrying out work in data analysis.

Deviation from the original plan of action

If considering when this research project was first proposed by Dr Bulgarelli, then it has changed quite substantially. However, that is due to the large time gap between when this project was proposed and when it began and thus, is to be expected.

Yet, the project has deviated from what was planned at the start. Due to the Covid-19 pandemic research labs throughout the UK were shut down and by the time they reopened my parallel Lab-based approach had to be scrapped. This resulted in the emphasis being placed heavily on the data analysis project and I was given new goals to further reinforce this research.

Was the experimental approach appropriate for my hypothesis?

Yes, my original hypothesis was that microbiome composition differs according to crop productivity, and if it does, to identify bacterial families significantly linked to differing crop productivity. This was accomplished and looking back it very likely could have been done with DESeq2 alone, however, as I mentioned earlier, the lack of a lab-based approach resulted in the addition of new goals for the data analysis approach, which were also accomplished with the approach utilized.

2. *List the most significant achievements of this project.* ***[400 words]***

The two most important results of this research are:

The identification of the Field effect.

A confirmation of the field effect was critical for the continuation of the rest of the project. Further dissertation goals were built on this being an affirmative result and as such was vital for the rest of the research.

Identification of bacterial families and genera linked with crop productivity

Although not as important as the first result, this finding added flavour to the rest of the project and allowed for speculation on what is happening in the spinach below-ground microbiome of the two fields. That speculation resulted in clues for where future research on this topic should focus on further developments in spinach microbiome exploitation.

3. *Did you meet your objective(s)? Explain your reasoning.* ***[500 words]***

Yes, I will discuss each objective, the strength of my approach, and the weakness below.

Objective 1: Identification of the Field effect.

My approach to accomplishing this was to use the ecological indexes Alpha and Beta diversity to visualize and test differences in the microbiomes of the two fields. The strength of this approach was that is speedily and easily accomplished this task along with generating informative figures to help a non-expert understand. However, the major drawback of looking for a field effect with only the ecological indexes is that I am unable to make any conclusive statement on whether the differences in microbiomes is because the Dragoni field has microbes that are better PGPRs, or because the Dragoni cultivar is more efficient at recruiting/recruits better PGPRs. To address this, I have proposed an approach in **Section 4.5**.

Objective 2: Identification of bacterial families and genera linked to crop productivity

My approach for this objective was to use the DESeq2 package to achieve it. The strength of a DESeq analysis is that it can clearly tell you what bacteria are differentially enriched between the two fields; however, it cannot offer any insights into the functionality of those differentially enriched bacteria. To account for that I did a mini-review of the literature

surrounding some of the more abundant bacterial families in **Sections 4.2**, **4.3** and **4.4**. Additionally, I again proposed another approach to address this issue in **Section 4.5**.

Section 4

4. *What was the most surprising thing you learned from this project?* ***[150 words]***

That the Phyloseq object essentially functions as a 4-dimensional dataset. It can store multiple datasets regarding different information that is still relevant to the overall structure. I am aware that python can do a very similar thing in something called ‘Dictionaries’ and ‘Arrays’ but was not at all aware of this functionality in R.

I, as of yet, have no clue how R accomplishes this, but I think it’s something do with the S3 and S4 class of objects in R studio and look forward to exploring this once time permits.

5. *How would you advise the next student to build on your discoveries?* ***[150 words]***

I would encourage anyone following up on this to look at **Sections 4.2**, **4.3**, **4.4**, **4.5** and **5** of this report as they detail, almost exactly, the approaches that can be used to further improve this research by addressing the shortcomings I did not have the time nor the global environment to attend to.

Appendix 3: Risk Assessment

RISK ASSESSMENT Date: 08/03/2020

1. **List activities or experimental procedures which may present significant risk to the health and safety of you and those in your working environment.**

No major hazards or risks, this work was carried out in silico.

2. **Major Hazards & Risks** (Be realistic; focus on the significant hazards & risks; discount the far-fetched & trivial; consider extent/routes of exposure; think about how likely it is that something might go wrong and what the outcome could be in terms of nature & severity of the harm/damage sustained)

Very little to none, possibly losing all my data due to some sort of electronical accident.

3. **List Specific Hazardous Substances, Items or Activities**

None, possibly RSI or back injury from sitting at the computer for too long

4. **Control Measures** (Briefly explain how the hazards identified in 3, above, above might be contained or mitigated)

Repeated breaks, proper posture, ergonomic chair, mouse and keyboard.

5. **Emergency Procedures** In addition to standard medical emergency, first aid, fire and spill procedures, identify if there are any additional special procedures which may apply:

No procedures apply to this project.

Waste Disposal Procedures

Does your project generate significant forms of waste material? **No**

If Yes:

In addition to standard waste disposal procedures (drain, waste bin) identify if there are special procedures for the disposal of specific waste items.

N/A