

Final Undergraduate Project

Bachelor in Biology

Amplicon metagenomics for elucidating microbial community
changes associated to conventional and organic farming
of *Beta vulgaris* and *Solanum lycopersicum*

Author

Olga Ibáñez Solé

Director

Miren Andone Estonba Rekalde

Co-Director

Iratxe Zarraonandia Martínez



Table of contents

1- ABSTRACT	1
2- INTRODUCTION	1
2.1- 16s rRNA amplicon metagenomics for studying plant-associated bacterial communities	1
2.2- Chloroplast and mitochondrial DNA interference when using EMP primers for studying plant-associated microbiomes	4
2.2.1- Post-sequencing chloroplast-sequence removal using bioinformatic tools	4
2.2.2- Post-PCR removal of chloroplast sequences from the amplicon set	4
2.2.3- Avoiding amplification of chloroplast sequences in the Polymerase Chain Reaction	5
2.2.4- Pre-PCR approaches for minimising plant-DNA in the DNA extracts	5
2.3- OBJECTIVES	7
3- MATERIALS AND METHODS	7
3.1 Organic and conventional soils	7
3.1.1- Sampling	7
3.1.2- Soil samples DNA extraction, PCR-amplification and Illumina-sequencing	8
3.1.3- Data analysis related to soil samples	8
3.2- Chloroplast interference assessment in aerial tissues	9
3.2.1- Aerial tissues sample collection	10
3.2.2- Sample processing before DNA extraction	10
3.2.3- DNA extraction	11
3.2.4- Measuring chloroplast interference	11
4- RESULTS	13
4.1- Organic and conventional agricultural soils	13
4.2- Alternative method for chloroplast interference assessment	17
4.3- Chloroplast interference assessment	17
5- DISCUSSION	18
5.1- Organic and conventional soil microbiomes	18
5.2- Alternative method for chloroplast interference assessment	20
5.3 – Chloroplast interference assessment	20
6- CONCLUSIONS	21
7- ACKNOWLEDGEMENTS	21
8- BIBLIOGRAPHY	21

1- ABSTRACT

Soil microbial community changes associated to conventional and organic farming of two relevant crops (*Beta vulgaris* and *Solanum lycopersicum*) were analysed through 16s rRNA amplicon sequencing. This study revealed microbial communities in the agricultural soils studied to be similar to other reported nutrient-rich microbiomes, and some significant differences between the microbial communities associated to the two farming practices were found. Some phyla (Chloroflexi and Thermi) were found to be present in different abundances according to soil treatment. As chloroplast interference can be a stumbling block in plant-associated 16s rRNA amplicon metagenomics analysis of aerial plant tissues, two protocols for bacterial cell detachment (orbital shaking and ultrasound treatment) and two protocols for microbial biomass recovery (centrifugation and filtration) were tested regarding their efficiency at excluding plant-DNA. An alternative method to the one proposed by Rastogi et al (2010) for evaluating the chloroplast-amplicon content in post-PCR samples was tested, and the method revealed that filtration was the most efficient protocol in minimising chloroplast interference.

2- INTRODUCTION

2.1- 16s rRNA amplicon metagenomics for studying plant-associated bacterial communities

There is an increasing debate in society on the hypothetical benefits of consuming organic farming products. Elucidating the physical and chemical differences as well as the differences regarding microbial communities associated to organic and conventional agricultural soils is of major importance and might shed some light on whether crops produced by conventional and organic farming systems are indeed different. Moreover, gaining knowledge related to organic and conventional plant-associated microbes could lead to more sustainable agricultural practices, as plant-microbe beneficial interactions can be exploited for minimising damage caused by crop pathogens or increasing crop yields (Martins et al, 2012).

The microbial composition of *Solanum lycopersicum* and *Beta vulgaris* has been previously studied by surveying their anatomical microbial ecology. Such work concluded that there are major differences in the microbial communities in different parts of the plant such as leaves, stem, fruits and roots and the plant surroundings such as in the rhizosphere and soil (Ottensen et al, 2013). Both endophytes (microorganisms living inside plant tissues) and epiphytes (microorganisms living on the surface of plant-tissues) have been studied. While the effect of farming systems on the microbial community associated to some plant species has been addressed (e.g. grape berries (Martins et al, 2012) and wheat (Esperschütz et al, 2007; Hartmann et al, 2006)), little is known about the impact of conventional and organic farming practices in the microbial communities associated to tomato (*Solanum lycopersicum*) and chard (*Beta vulgaris*). That is why the present

study pretends to fill that gap of knowledge by studying microbial communities in soil and different plant tissues (root, leaves and fruits) of both species grown under the two aforementioned farming practices.

There are two main approaches for studying microbial communities: 1) culture-based approaches (the traditional method) and 2) culture-independent methods.

In the culture-based one, samples are inoculated in culture media, the Colony Forming Units (CFUs) are counted and further biochemical analyses are performed in order to characterise those CFUs. Then, quantitative aspects of the community can be inferred to some extent. Since most microorganisms seem to be unculturable, this approach is rather limited in order to describe the taxonomic structure of microbial communities. In fact, new strategies have shown that only about 1% of the microorganisms are culturable (Schloss and Handelsman, 2003). In other words, culture-based approaches select for culturable microorganisms while ignoring non-culturable ones, leading to incomplete or biased community diversity assessments.

In culture-independent methods (also called direct methods), DNA is extracted and sequenced directly from environmental samples and communities are analysed by comparing the sequence-composition of the sample. Culture-independent methods include shotgun metagenomics (also known as metagenomics) and amplicon metagenomics (or amplicon sequencing) (Wooley et al, 2010).

In shotgun metagenomics, the total genomic DNA of an environmental sample is fragmented and sequenced using next generation sequencing technologies. Then, sequences are assembled to obtain genome contigs and those are sorted in groups that might represent an individual genome or genomes of closely related organisms in a step called binning. Finally, metagenomes are interpreted in the annotation step (Thomas et al, 2012). The shotgun approach looks for the potential functions in the community and thus, it answers to “what the microorganisms could potentially do”. This method can lead to the discovery of new enzymes, unravelling function-phylogeny links or characterizing evolutionary profiles of microbial communities (Thomas et al, 2012).

Conversely, amplicon metagenomics are phylogenetic surveys based on the diversity of a single gene (a phylogenetic marker) (Thomas et al, 2012), and so they answer to answers to “which microorganisms are there”. In this approach, a phylogenetic marker (e. g. 16s rRNA gene for studying bacteria, 18s rRNA gene for eukaryotes, internal transcribed spacer (ITS) gene for fungi) is amplified and sequenced and the differences on the sequence of these polymorphic markers allow researchers to classify them into taxa. Therefore, the diversity of different taxa in the community can be inferred.

The 16s ribosomal RNA (rRNA) gene is considered to be a well-suited marker for amplicon metagenomics phylogenetic surveys that aim to study the bacterial community, as this gene shows enough polymorphism to enable us to discern between different taxonomic groups, but it is not too polymorphic, so similarities between sequences coming from related taxa can be identified (Klinworth et al, 2012). However, some authors argue that 16s rRNA might not be as convenient of a marker since it has variable copy numbers, with some taxa having up to 15 copies of this gene (Kembe et al, 2012). As a consequence, those taxa might be overrepresented in the final set of sequences and that could give rise to biased community structure estimations. Even so, the advantages of using 16s rRNA genes outweigh this inconvenience and that is why it is widely used in amplicon sequencing surveys. Primers for 16s rRNA amplification are designed so that they bind to the conserved regions on both sides of a variable region, generating amplicons of the polymorphic regions which can be taxonomically classified according to their sequence (Figure 1).



Figure 1 Conserved (green) and variable (V1-V9, grey) regions in 16s rRNA genes. Source: Alimetrics (<http://www.alimetrics.net>).

The choice of primers is believed to be the most critical step for accurate rDNA amplicon analysis, as using suboptimal primers can lead to the under-representation or selection against some particular species or even whole groups. For instance, Klindworth et al (2012) evaluated the overall coverage and phylum spectrum of the most widely used primer pairs. In that study, the primer pair *515f/806r* (used in the Earth Microbiome Project) was shown to be the one of broadest spectrum.

The Earth Microbiome project (EMP, <http://www.earthmicrobiome.org>) seeks to create a database of comparable microbiomic data from different environments across the globe (Gilbert et al, 2014). For that purpose, a standard protocol and pipeline was created to be applied to all types of samples (same primers, similar DNA extraction protocol, etc.) in order to make all the results comparable. It was crucial to find a primer pair to amplify a variable region in the 16s rRNA gene that would be informative for many different taxa. Primers *515f/806r* (hereafter: “EMP primers”) were chosen because of their capacity for evenly amplifying sequences from a broad range of taxa. The present study is embedded in such big microbiome effort and thus, it follows the EMP standards and uses the primers described.

2.2- Chloroplast and mitochondrial DNA interference when using EMP primers for studying plant-associated microbiomes

One of the main obstacles when using 16s amplicon metagenomics for evaluating the microbial diversity of plant-associated communities is that most primers (including EMP primers) amplify eukaryotic organelle DNA as well. As a consequence, many of the sequences we will obtain will correspond to mitochondrial and chloroplast DNA sequences from the plant genome, meaning that most of the sequencing power will be used for reading sequences of no interest for our study. This is especially true for studies that aim to analyse endophytic microbial diversity, as DNA extraction is performed directly from plant tissues, increasing the plant-DNA content in the extract. For instance, a research project on *Vitis vinifera* bacteria that used EMP primers showed that chloroplast sequences posed the biggest problem, since they can make up to 98% of the sequences obtained using the Illumina MiSeq sequencer. In contrast, mitochondrial sequences do not seem to be such big of a problem (Zarraonaindia et al, 2015).

In the present study, where we aimed to study the microbial communities of several plant tissues such as stem, leaves and fruits using EMP primers, strategies were needed in order to minimise plastid sequences, and particularly chloroplast sequences, in our dataset.

Following the standard workflow of a 16S rRNA amplicon metagenomics project, there are in theory four main stages at which we can act against the overrepresentation of plastid sequences: post-sequencing (using bioinformatic tools), post-PCR (using common molecular biology techniques), in the PCR amplification step (playing with different sets of primers) and pre-PCR (during sample processing and DNA extraction).

2.2.1- Post-sequencing chloroplast-sequence removal using bioinformatic tools

Sequences coming from plant organelles can be removed by means of bioinformatic tools in the data analysis step. Even though it is a fairly easy solution, it has some important disadvantages. First, it is done in a post-sequencing stage, and thus, it has no effect on the loss of sequencing efficiency, usually rendering it not cost-effective. In addition, having too many chloroplast sequences (and therefore too few bacterial sequences) could bias bacterial population-size estimations.

2.2.2- Post-PCR removal of chloroplast sequences from the amplicon set

Chloroplast sequences can be separated from target sequences after the PCR in case they show some difference in length that enables us to discern between them using common molecular biology techniques.

If bacteria and chloroplast amplicon-sizes are enough different, unwanted sequences can be removed by gel electrophoresis. EMP primers generate similar sized bacterial and chloroplast amplicons, making this approach impracticable. Another disadvantage is that extracting the target band from a gel might decrease

the quality of the PCR product, and that is likely to make sequencing more complicated.

In case chloroplast sequences have a single target for a restriction enzyme that is absent in bacterial amplicons, they can be digested so that we obtain different-sized DNA fragments that can be separated by gel electrophoresis (e.g. two bands coming from the specific digestion of chloroplast-amplicons and another band corresponding to bacterial undigested amplicons). However, in order to use this approach, researchers must ensure that none of the bacterial amplicons have a target for the restriction enzyme, which might not be easy, because many of the microorganisms and their 16s rRNA gene sequences are unknown before the sequencing step.

2.2.3- Avoiding amplification of chloroplast sequences in the Polymerase Chain Reaction

There are two main ways by which PCR-amplification of unwanted chloroplast sequences can be avoided. The most straight-forward one is to use primers that specifically exclude chloroplast sequences. Several pairs of primers have been tested, showing different levels of success at excluding unwanted chloroplast sequences (reviewed in Hanshew et al 2013) and variable reliability to amplify bacterial sequences.

Additionally, Lundberg et al (2013) developed blocking primers called peptide nucleic acid (PNA) PCR clamps which bind to chloroplast amplicons and specifically inhibit their further amplification. The latter method appears as a very promising strategy, even though it is somewhat expensive and it is not widely used still.

2.2.4- Pre-PCR approaches for minimising plant-DNA in the DNA extracts

The process prior to DNA extraction is crucial in the effort of minimising chloroplast interference. When analysing endophytes, chloroplast sequences cannot be avoided, because plant tissues are used as the input for DNA extraction and plant-DNA and microbial DNA are mixed in that process. Conversely, epiphyte studies aim to select for microbial communities present on plant surfaces and, as a consequence of sample processing, plant sequences might be excluded.

A thorough revision on the commonly used protocols for sample processing in epiphyte community studies showed that the common processing consists of two steps: first, microbial cells are detached from plant surfaces and then, those cells are isolated so that microbial DNA can be extracted from them. Different protocols have been described for the first detachment step, including either swirling plant tissues (Martins et al, 2013; Rastogi et al, 2010) or subjecting them to an ultrasound treatment (Bulgarelli et al, 2012) in a buffered solution. Both treatments result in a buffered bacterial cell suspension, from which microbial biomass can be isolated in two ways: either centrifugation (Rastogi et al, 2010) or filtration (Martins et al, 2013) (Figure 2).

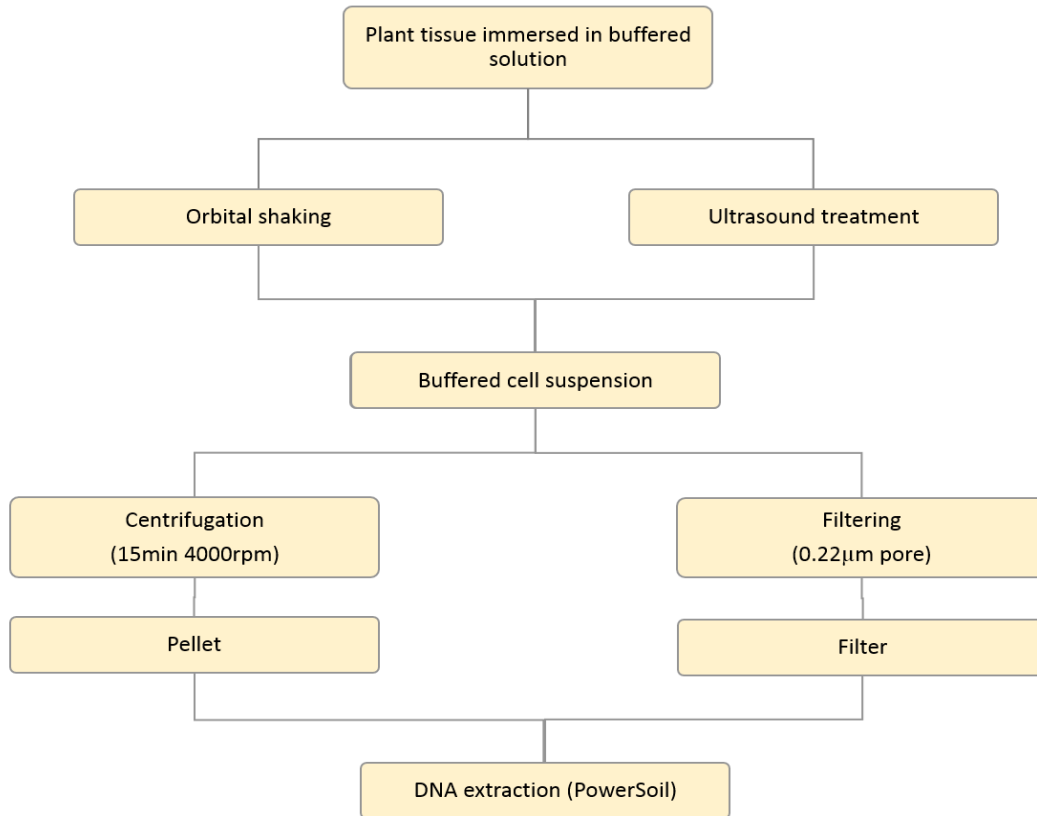


Figure 2 General workflow for sample processing prior to DNA extraction in epiphyte studies, considering the most frequently used protocols for bacterial cell detachment and recovery.

Studying the effect of using either one or other combination of protocols in the resulting chloroplast contamination would be useful in order to minimise the chloroplast content prior to the extraction step and as a consequence in the subsequent sequencing. While authors used either one or other protocol, there is not any published work to the best of our knowledge comparing the efficiency of such protocols. For that reason, in the present study we designed and assay in which comparable samples of *Beta vulgaris* leaves would be processed using every possible combination of the protocols to be studied (orbital shaking, ultrasound treatment, filtration and centrifugation). Although sequencing those samples' 16s rRNA PCR products would be the most direct way for evaluating which protocol was the most effective in minimising chloroplast sequences (without having an effect on the overall community structure, this would be very expensive, and so in the present study we tried to find an alternative protocol.

2.3- OBJECTIVES

The present work's aims are:

- 1) To compare the taxonomic structure of conventional and organic agricultural soils associated to chard and tomato plant species, to see whether there is a significant difference regarding bacterial composition and diversity between the two farming systems.
- 2) To study the effect of using different combinations of protocols during the processing of aerial plant-tissue samples prior to DNA extraction in the resulting chloroplast interference. In particular, evaluating two protocols for bacterial cell detachment from plant surfaces (orbital shaking and ultrasound treatment) and two protocols for bacterial cell recovery from the cell suspension (centrifugation and filtration).

3- MATERIALS AND METHODS

3.1 Organic and conventional soils

3.1.1- Sampling

Two 10 m²-sized land plots (35m away from each other) located in Beotegi (43° 5.370' N; 3° 4.590' W) were treated using two different farming techniques: conventional and organic. In the conventional plot, 0.25 kg·m⁻² of a chemical fertilizer (NPK 15.15.15 (15); Fertiberia, S.A.) were applied once 25 days before the plantation. As a result, a dose of 375 g of N, P₂O₅ and K₂O was supplied to the conventional plot. Phytosanitary treatments included a liquid mixture of an insecticide (Epik20SG; Sipcam Jardín S.L.) and a fungicide (Galben M.; Sipcam Jardín S.L.) applied twice (200 ml m⁻² in total), 7 and 36 days after plantation. In the organic plot, a natural fertilizer (natural horse manure, Abonos Naturales Hermanos Aguado, S.L.; product approved and certified by CAEE as ecological product; C qualification) was spread in two phases, 10 and 2 days before the plantation, using a total of 6.48 kg m⁻², which corresponded to supplied doses of 2981 g of N, 259 g of P₂O₅ and 583 g of K₂O. Protective plants (*Tagetes patula*) were planted in the periphery of the organic plot as natural repellent to avoid pests and insect attacks.

Twenty-five chard (*Beta vulgaris*) and tomato (*Solanum lycopersicum*) seedlings were transplanted in each plot in June 2013 and grown until October 2013. Three sampling campaigns were performed in June, July and August 2013, in which soil samples from the root-zone of the two plant species as well as plant-tissue samples (root, stem, fruits and leaves) were collected in each plot. Samples were carried to the laboratory and stored at -20°C until processing.



3.1.2- Soil samples DNA extraction, PCR-amplification and Illumina-sequencing

0.25g of soil were used as an input for DNA extraction using PowerSoil DNA extraction kit (MoBio). Genomic DNA was amplified using the EMP barcoded primer set, adapted for Illumina MiSeq by adding nine extra bases in the adapter region of the forward amplification primer that support paired-end sequencing. The V4 region of the 16S rRNA gene (*515f/806r*) was amplified with region-specific primers that included the Illumina flowcell adapter sequences. The reverse amplification primer also contained a twelve base barcode sequence that supports pooling of up to 2,167 different samples in each lane. Each 25µl PCR reaction contains 12µl of MoBio PCR Water (Certified DNA-Free), 10µl 5µM HotMasterMix (5 Prime), 1µl of Forward Primer (5µM concentration, 200pM final), 1µl Golay Barcode Tagged Reverse Primer (5µM concentration, 200pM final), and 1µl of template DNA. The conditions for PCR were as follows: 94°C for 3 minutes, with 35 cycles at 94 °C for 45 s, 50 °C for 60 s, and 72 °C for 90 s; with a final extension of 10 min at 72 °C. PCR amplifications were then pooled. Following pooling, amplicons were quantified using PicoGreen (Invitrogen) and a plate reader. Once quantified, different volumes of each of the products were pooled into a single tube so that each sample was represented equally. This pool was then cleaned using the UltraClean® PCR Clean-Up Kit (MoBio) and quantified using Qubit (Invitrogen). After quantification, the molarity of the pool was determined and diluted to 2nM, denatured for sequencing on the Illumina MiSeq platform (150bp x 2 pair end) at Argonne National Laboratory (head group of the EMP project).

3.1.3- Data analysis related to soil samples

Sequences were obtained from the Illumina sequencing reaction in a Fastq file that contained, in addition to the nucleotide sequences, their corresponding quality scores. Data analysis was performed using Qiime version 1.8 (Caporaso et al, 2010). First, *split_libraries_fastq.py* command was used in order to link each sequence to the sample it came from. This command demultiplexes the sequences using the Forward and Reverse fastq sequence data files and the barcode-sequence files and fusing the information into a single fasta file. In addition, this script performs a quality trimming step whereby sequences that are too short or sequences containing ambiguities are discarded from the dataset.

An OTU table was created using the script *pick_open_reference_otus.py*. This script clusters our sequences into OTUS (operational taxonomic units) by sequence similarity (97% nucleotide similarity) and compares those OTUs with all of the 16s rRNA gene sequences available in the *Greengenes* database to assign them a taxonomy. Then, a subsampled OTU table was created using the script *single_rarefaction.py*, whereby the original OTU table was rarefied to 32 995 sequences per sample in order to avoid biases related to obtaining different sequencing depths per sample.



Alpha diversity metrics were computed and alpha rarefaction plots were created using the script [alpha_rarefaction.py](#), considering the samples first as a single set and then sorted by treatment (conventional and organic). Beta diversity and 3D PCoA plots were obtained using the script [beta_diversity_through_plots.py](#). One of the samples (corresponding to *Beta vulgaris* organic agricultural soil) was filtered out from the OTU table since it was inconsistent with the rest of the samples. This was done using the script [filter_samples_from_otu_table.py](#).

New OTU tables sorted by treatment (conventional and organic) and by variety (*Solanum lycopersicum* and *Beta vulgaris*) were created using the script [split_otu_table.py](#). Then beta diversity was obtained for each subset ([beta_diversity_through_plots.py](#)).

In order to test whether there were significant differences between organic and conventional communities within tomato and chard categories, we used the script [compare_categories.py](#) to conduct Adonis and ANOSIM tests. Adonis is a nonparametric statistical method that takes a beta diversity distance matrix, a mapping file and a category in the mapping file to determine sample grouping from. In this study the weighted and unweighted Unifrac distances were used and the effects of the category “treatment” (conventional/organic) and category “variety” (tomato/chard) were tested. Adonis computes an R^2 value (effect size, which shows the percentage of variation explained by a category) as well as a p-value to determine the statistical significance. In the ANOSIM test, the dataset is separated into groups according to a category in the metadata mapping file (here “treatment” and “variety”, respectively) and it tests whether there are significant differences between those groups. Both tests’ significance, ANOSIM and Adonis, was determined through permutations (1000).

The script [group_significance.py](#) was used for comparing phylum mean frequencies in sample groups and to ascertain whether or not there were statistically significant differences (according to the Kruskal-Wallis test) in their abundances between the different sample categories (treatment and variety).

3.2- Chloroplast interference assessment in aerial tissues

While it was the initial aim to study the microbial communities in chard and tomato tissues (stem, leaves and fruits), the co-amplification of chloroplast-DNA together with the bacterial 16s rRNA gene precluded us to do so in the present study due to the uncertainty of which would be the best protocol for processing those samples. Thus, despite those samples were collected and kept frozen from the two Beotegi plots previously treated with the two different farming practices, we did not process such samples. Besides, it was necessary to first design an experiment that would enable us to test the most efficient protocol in minimising chloroplast sequences.

3.2.1- Aerial tissues sample collection

In order to determine the best protocol for minimising chloroplast interference among the four protocols explained above, two *Beta vulgaris* whole plants were taken from a greenhouse in Igeldo mountain (San Sebastian), in March 2015. It was not possible to collect tomato samples as it was not the growing season for this species. *Beta vulgaris* samples were carried to the laboratory in a plastic bag. Eight leaves were cut off, placed in sterile individual plastic bags and stored at 4°C until processing.

3.2.2- Sample processing before DNA extraction

a) Bacterial cell detachment from chard surfaces was done in two ways:

Orbital shaking: for each sample, 200 mL of PBS were introduced in a plastic bag containing a single *Beta vulgaris* leaf. Plastic bags were placed on an orbital shaker at maximum speed for 1 hour.

Ultrasound treatment: 200 mL of PBS were introduced in each plastic bag, then plastic bags were put one by one into a sonication bath for 4 minutes. Special care was taken not to keep any samples immersed in PBS longer than others. Samples were kept ice-cold while waiting.

b) Cell isolation from the cell-suspension obtained in the previous cell detachment step was done in two ways:

Centrifugation: for each sample, about 50 mL were introduced in sterile nucleic acid-free Falcon tubes and were centrifuged for 15 minutes at 4000 rpm (4°C). The supernatant was discarded and each tube was filled up again with cell-suspension. This step was repeated 4 times, until no liquid was left inside each of the bags. Finally, pellets were re-suspended using approximately 20 mL of cold PBS and centrifuged for 25 minutes at 4000 rpm. The supernatant was discarded and tubes were stored at -20°C.

Filtration: the resulting 200 mL of cell-suspension were filtered using first 8 µm pore-sized filters to get rid of plant cell debris (since during the trial period we had observed that 0.22 µm filters soon became saturated with debris in unfiltered cell-suspensions). Then, the flow-through was collected and transferred to a sterile glass beaker. The collector vessel was rinsed with 50 mL of PBS that were then transferred to the same beaker. The collector vessel was washed using soap and water, then rinsed with bleach and finally with distilled water. The flow-through was re-filtered using a 0.22 µm pore-sized filter. In order not to lose many bacterial cells, the glass beaker was also rinsed using 50 mL of PBS that were then re-filtered. The 0.22 µm filter containing bacterial cells was picked using sterile Millipore stainless steel filter holders and stored in a sterile tube at -20°C.

The combination of the two protocols for bacterial cell detachment and two methods for bacterial recovery resulted on four protocols for sample processing to be tested: 1) Orbital shaking + centrifugation, 2) Orbital shaking + filtration, 3) Ultrasound treatment + centrifugation, and 4) Ultrasound treatment + filtration. Each combination was tested with two replica of *Beta vulgaris* leaves.

3.2.3- DNA extraction

From the four samples that had been processed following the centrifugation protocol, DNA was extracted directly from the pellet obtained using the PowerSoil DNA Extraction kit (MoBio). In the case of the four samples that had been processed following the filtration protocol, half of each of the 0.22µm filters was used as an input for DNA extraction using PowerSoil DNA extraction protocol (MoBio), following PowerSoil kit instructions, with the addition of 60µL of Tris/EDTA/SDS buffer in the first step (as it had shown to be optimum in previous trials). DNA was eluted using 60µL of elution buffer (instead of 100µL as in the original PowerSoil kit protocol). DNA concentrations in the extracts were quantified using a NanoDrop UV -spectrophotometer.

3.2.4- Measuring chloroplast interference

For estimating chloroplast interference, a modified version of the method proposed by Rastogi et al (2010) was used. Their method consisted on amplifying DNA extracted from plant surface wash samples using the primers *27f-YM* and *Eub518r* to obtain ~500bp-long amplicons, where the chloroplast-amplicon of lettuce they analysed is slightly smaller (471bp) and has a target for the restriction endonuclease *BbvCI* that is absent in bacterial amplicons. Even though this method is a powerful tool that can be used in a number of plant species (e.g. *Vitis vinifera* and *Solanum lycopersicum*) for estimating the chloroplast-amplicon content of a sample after PCR without the need of sequencing, it is not directly useful for *Beta vulgaris*, as this species' chloroplast-amplicon lacks a target for *BbvCI* (Figure 3 and Table 1).

<i>E. coli</i>	5'-AGAGTTTGATCATGGCTCAG.....CCAGCAGCCGCGGTAAT -3'
<i>P. aeruginosa</i>	5'-AGAGTTTGATCATGGCTCAG.....CCAGCAGCCGCGGTAAT -3'
<i>V. vinifera</i>	5'-AGAGTTTCGATCCTGGCTCAG.....GCTGAGG.....CCAGCAGCCGCGGTAAT -3'
<i>S. lycopersicum</i>	5'-AGAGTTTCGATCCTGGCTCAG.....GCTGAGG.....CCAGCAGCCGCGGTAAT -3'
<i>B. vulgaris</i>	5'-AGAGTTTCGATCCTGGCTCAG.....GCTGAGA.....CCAGCAGCCGCGGTAAG -3'

Figure 3 Restriction site (green) for endonuclease *BbvCI* in different plant species and two examples of bacteria (*Escherichia coli* and *Pseudomonas aeruginosa*) showing there is a mismatch (red) in *Beta vulgaris*.

Species	Amplicon size (bp)	Fragment length after <i>BbvCI</i> digestion (bp)
<i>Escherichia coli</i>	527	527*
<i>Pseudomonas aeruginosa</i>	521	521*
<i>Vitis vinifera</i> chloroplast	474	160 + 314
<i>Solanum lycopersicum</i> chloroplast	483	169 + 314
<i>Beta vulgaris</i> chloroplast	474	474*

Table 1 expected amplicon sizes and the expected fragment sizes after *BbvCI* digestion for plant and two bacteria species. (*) Amplicons that lack the target sequence for *BbvCI*.

In order to evaluate the four protocols for pre-PCR processing in the particular case of *Beta vulgaris* we needed to use an alternative protocol to the one described by Rastogi et al (2010). The method we used is based on the hypothesis that the lower band resulting from the gel electrophoresis of *27f-YM/Eub518* primers amplification belongs exclusively to plant chloroplast. If that were the case, the intensity of the fluorescence emitted by the upper (bacterial) and the lower (chloroplast) band in each lane could be measured and that would serve as a proxy to evaluate the method that yielded fewer chloroplast amplicons. In order to use this alternative method, it was necessary to sequence the lower band to ensure it belonged exclusively to chloroplast sequences.

3.2.4.1- Testing the alternative method for chloroplast interference assessment by Sanger sequencing

Direct DNA extracts from *Beta vulgaris* leaves and *Vitis vinifera* grapes were used as chloroplast-positive controls. A soil sample was used as negative-control as we expected chloroplast sequences to be neglectable in such sample. Two *Beta vulgaris* samples (a frozen and a fresh leaf sample) processed following the previously described orbital shaking + centrifugation protocol were tested.

Four PCR reactions were done with each of the DNA extracts using the primers *27f-YM* and *Eub518r*. The four PCR replica for each sample were mixed and run together on a 2% agarose gel until the two bands were properly separated. The lower band in each lane was cut using sterile scalpels and DNA was extracted from them using the QIAquick Gel Extraction kit. The DNA concentration of the extracts was measured using NanoDrop. The DNA extracted from the lower band of the two chloroplast-positive controls (*Beta vulgaris* leaf extract and *Vitis vinifera* grape extract), as well as the lower band in the frozen sample were Sanger-sequenced in SGIker.

3.2.4.2- Measuring relative chloroplast and bacterial amplicon content

Once the alternative method for the chloroplast interference assessment had proven to be accurate (the lower band belonged exclusively to chloroplast sequences), such method was used for comparing the chloroplast-amplicon content of the samples processed following the four protocols to be studied. The DNA extracts of the samples processed using the 4 protocols 1) Orbital shaking + centrifugation, 2) Orbital shaking + filtration, 3) Ultrasound treatment + centrifugation, and 4) Ultrasound treatment + filtration) were amplified using *27f-YM* and *Eub518r* primers following Rastogi et al (2010) PCR protocol (changing the annealing temperature to 55°C). Amplicon concentration of the purified PCR products were adjusted to 50 ng/μL. Then, 5 μL of each sample (250 ng of DNA) were loaded in a 2% agarose gel and run at 80 V for 40 minutes. The intensity of the microbial and the chloroplast band were measured in CHEMI and the INTm:INTc ratio (Intensity of the microbial band/Intensity of the chloroplast band) was calculated in each lane.

4- RESULTS

4.1- Organic and conventional agricultural soils

Considering all of the soil samples together, 1040 OTUs were found in total. One sample was filtered out from the OTU table for being considered an outlier and six other samples were discarded because they lacked the minimum amount of sequences (32995) specified as cut-off.

The most abundant phyla in soil communities were Proteobacteria (40%), Acidobacteria (%16), Actinobacteria (%12) and Bacteroidetes (%7) (Figure 4a). Those four phyla accounted for, on average, 75% of the total abundance in the community. The most abundant classes were Alphaproteobacteria (14%), Deltaproteobacteria (%10), Acidobacteria-6 (%10), Betaproteobacteria (%9) and Gammaproteobacteria (%7) (Figure 4b). Those five classes made up, on average, 50% of the soil community.

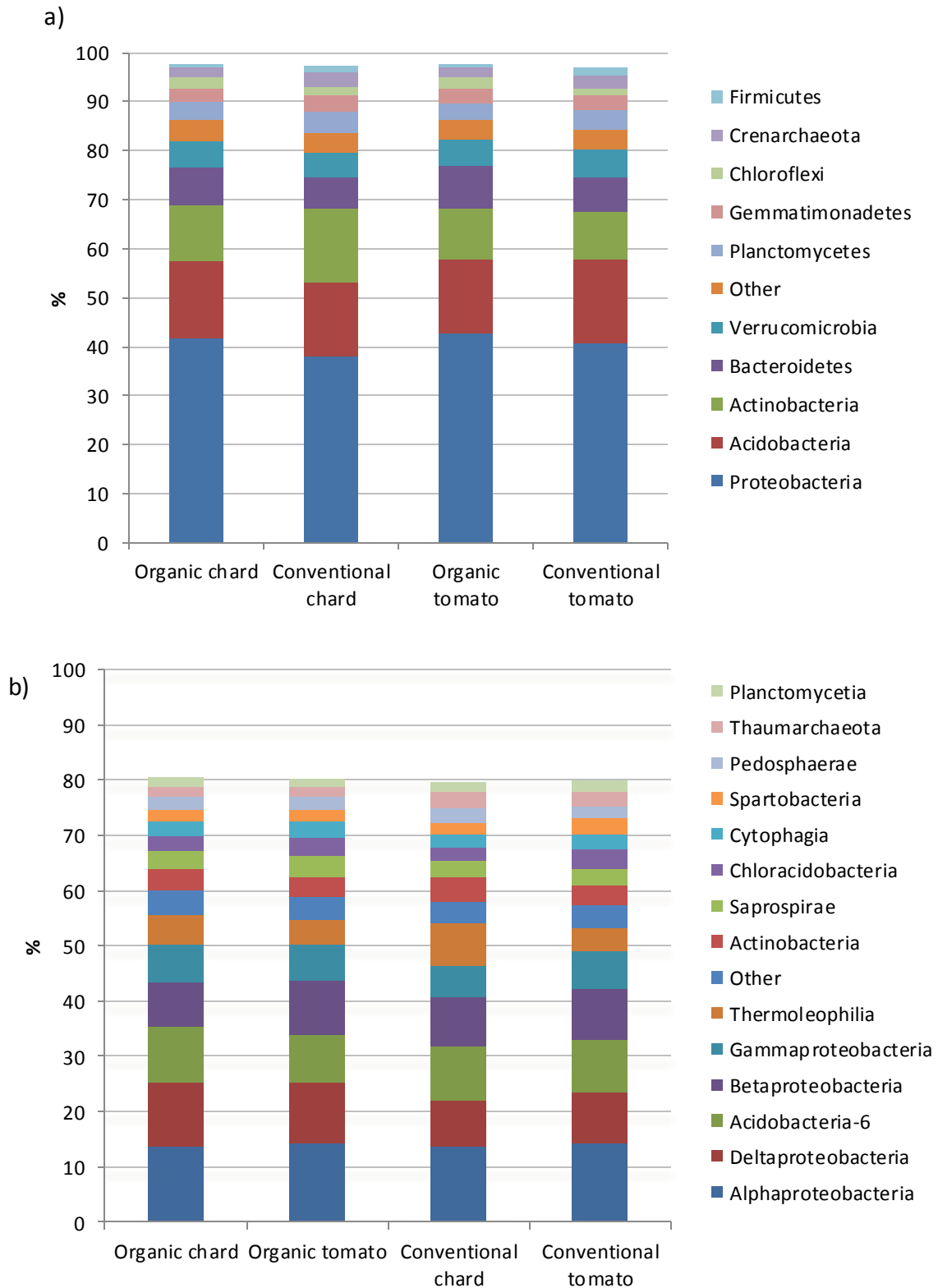


Figure 4 Soil microbial community structure by taxonomic rank. Most abundant a) phyla and b) classes in the communities are shown sorted by treatment (mean relative abundance of taxa in organic and conventional samples) and variety (mean relative abundance of taxa in chard and tomato). Only phyla and classes that represented more than the 2% of the total abundance are shown in the barchart.

Regarding alpha diversity, we could not see any significant difference in species richness between samples, neither among treatment nor variety categories (Figure 5).

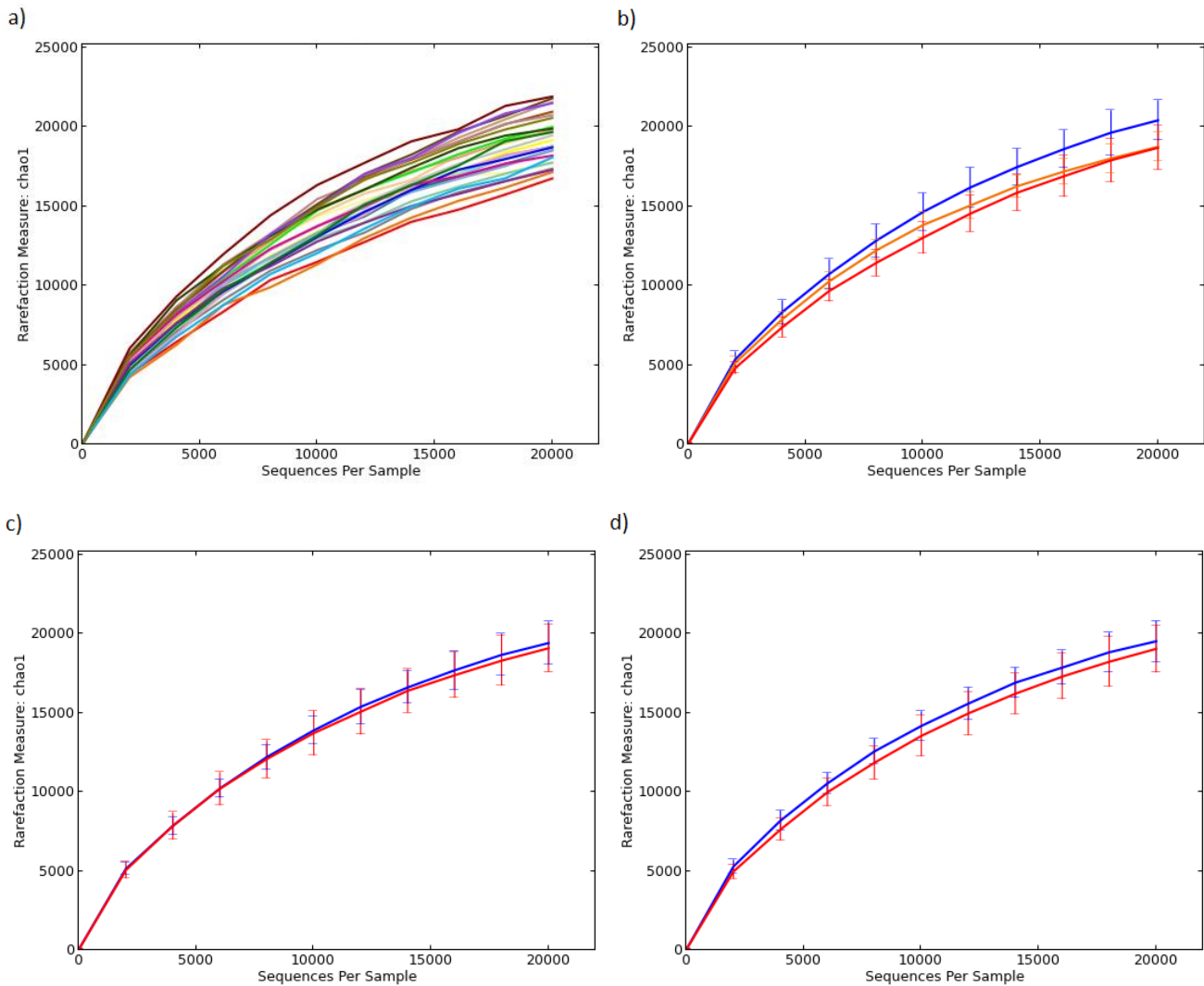


Figure 5 Alpha rarefaction plot of **a)** all samples, and mean rarefaction curves of samples sorted by **b)** month, **c)** treatment and **d)** variety.

When analysing differences between the soil communities studied (beta-diversity), the 3D PCoA plot for all samples (Figure 6) showed bacterial communities in organic and conventional samples to differ, despite the differences observed between the two farming systems represent only 6% of the total variability in the dataset. Therefore, the farming practice was the main variable shaping the microbial community structure, and neither the variety nor the collection month (temporal scale) showed to have a significant influence on the community.

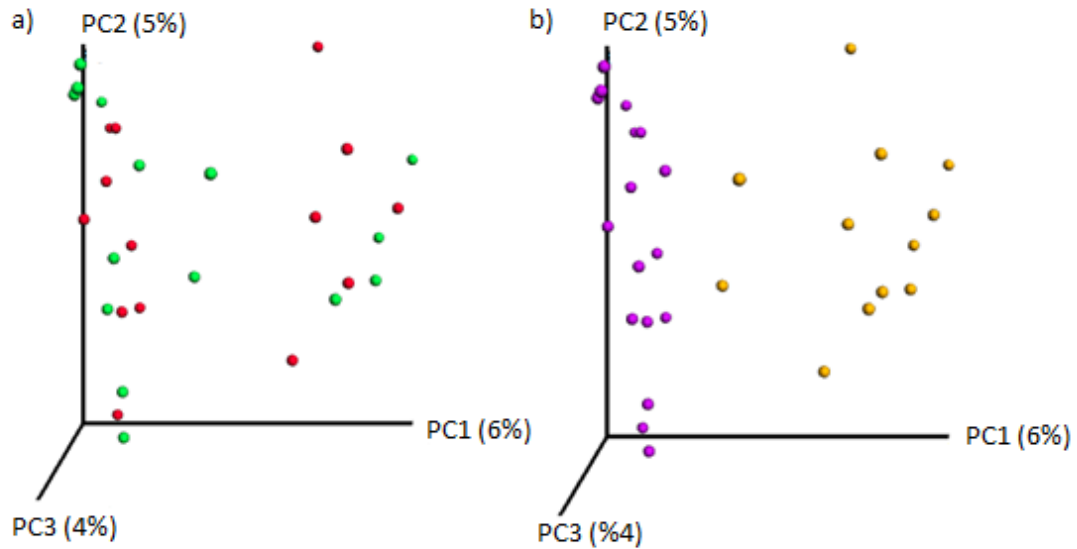


Figure 6 3D PCoA plot considering all soil samples. The same plot is shown with sample-points coloured according to **a)** variety (red = tomato, green = chard) and **b)** treatment (purple = conventional, yellow = organic).

Considering the effect of farming practices aside plant variety, we found significant differences in the taxonomic structure of the microbial communities in organic and conventional agricultural soils. The R^2 value (0.1380) in the Adonis test using weighted UniFrac distances shows that about 14% of the total variation in the dataset is explained by the conventional vs organic grouping (Table 2). The results obtained in the ANOSIM test evidence that there are significant differences between conventional and organic soil communities ($R=0.3283$, $p=0.01$) (Table 2).

UniFrac distance (conventional vs organic)	<i>Adonis</i>				<i>ANOSIM</i>			
	Unweighted		Weighted		Unweighted		Weighted	
	R^2	p	R^2	p	R	p	R	p
a) Tomato	0.1022	0.001	0.2106	0.002	0.7258	0.001	0.4990	0.01
b) Chard	0.09112	0.001	0.1524	0.04	0.5744	0.01	0.2193	0.03
c) All samples	0.05655	0.01	0.1380	0.01	0.06325	0.01	0.3283	0.01

Table 2 R^2 , R and p values obtained in the Adonis and ANOSIM tests for organic and conventional farming soil. The test shows differences related to farming practices **a)** within tomato farming soils, **b)** within chard farming soils and **c)** considering both tomato and chard farming soils together.

Considering that the crop variety might have an effect on the way microbial communities respond to either organic or conventional farming practices and that soils planted with different crops might show differences in their microbial communities despite being treated following the same agricultural practices, we had run Adonis and ANOSIM tests to compare the effect of soil treatment within tomato and chard farming sample groups.

The strongest effect of soil treatment was observed in tomato farming soils, where differences between conventional and organic communities made up %21 of the total variation (adonis test R^2 using UniFrac weighted distances). The ANOSIM test supported this result and showed an R value of 0.7258 ($p=0.001$) for this correlation.

Chard farming soils also showed differences according to treatment, but those differences were only significant ($p < 0.01$) when using unweighted distances.

Kruskal-Wallis test showed that Chloroflexi and Thermi were the only phyla to have significantly different abundances in organic and conventional soils (Bonferroni $p=0.0017$ and 0.005 respectively). Both were more abundant in organic than in conventional soils.

4.2- Alternative method for chloroplast interference assessment

The Sanger sequencing of the lower band of the chloroplast-positive controls (*Beta vulgaris* and *Vitis vinifera* samples where DNA was directly extracted from leaf tissue) resulted in a single sequence that coincides with those species' chloroplast sequences according to Blastn alignment using NCBI database.

Regarding *Beta vulgaris* samples processed with the orbital shaking + centrifugation protocol, both the concentration and the quality of the DNA extracted from the lower band from the fresh *Beta vulgaris* leaf wash sample was too low and it failed to emit a strong signal during Sanger sequencing. However, the Sanger sequencing of the lower band of the *Beta vulgaris* frozen leaf sample was successful, and resulted in a unique sequence that aligned with the *Beta vulgaris* chloroplast sequence in a ClustalX alignment.

4.3- Chloroplast interference assessment

We could not perceive any difference neither in the overall DNA yield (Table 3) nor in the relative amounts of bacterial and chloroplast amplicons related to the use of one or other bacterial cell detachment protocols (Ultrasound/Orbital shaking) (Figure 7 and Table 3). However, there was a difference related to the method used for cell recovery, as the centrifuging protocol showed to result in higher yields of total DNA than the filtering protocol. Nevertheless, the relative amount of chloroplast sequences was smaller in samples that had been processed following the filtration protocol.

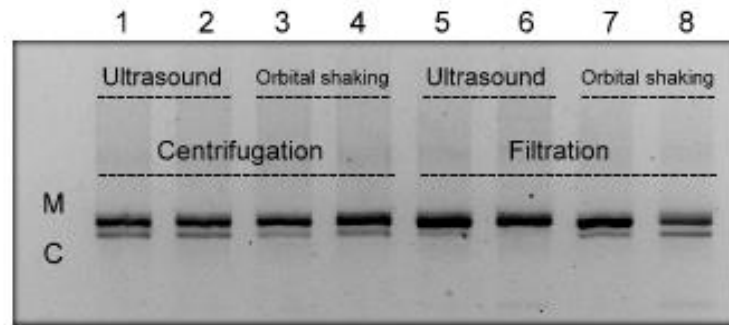


Figure 7 Electrophoresis gel with the purified 16s PCR products obtained from processing eight fresh *Beta vulgaris* leaves using four combinations of protocols: M= Microbial band; C = Chloroplast band; Lanes 1, 2, 5 and 6: ultrasound treatment; lanes 3, 4, 7 and 8: orbital shaking; lanes 1-4: centrifugation; lanes 5-8: filtration.

Detachment	Recovery	DNA (ng/ μ L)	INTm/INTc
Ultrasound treatment	Centrifugation	20.490	1.47
		16.960	1.42
Orbital shaking	Centrifugation	23.240	1.55
		19.220	1.57
Ultrasound treatment	Filtration	2.827	2.37
		4.061	2.30
Orbital shaking	Filtration	4.220	2.50
		5.285	3.85

Table 3 DNA concentration (ng/ μ L) and microbial to bacterial band intensity ratio (INTm/INTc) obtained in the PCR products of eight *Beta vulgaris* leaf wash DNA extracts that had been processed following four combinations of protocols for bacterial cell detachment and recovery (two replica per combination).

5- DISCUSSION

5.1- Organic and conventional soil microbiomes

The bulk soil samples analysed in the present study reflect a nutrient rich environment, as demonstrated by the ratio of Proteobacteria to Acidobacteria found, as it has been reported that Proteobacteria are favoured in nutrient rich soils, while Acidobacteria are more abundant in nutrient-poor soils (Fierer, Lauber et al, 2012; Fierer, Left et al, 2012).

In addition, the present study reveals that the soil microbiomes associated to the two farming systems (conventional and organic) are significantly different. In concordance, a three-year study that compared 24 conventional and organic agricultural soils revealed that the relative abundance of some of the less dominant phyla were affected by soil treatment, whereas changes in the abundances of larger phyla were related with sample year rather than conventional or organic agricultural practices. In particular, Chloroflexi and Nitrospirae were found to be significantly more abundant in organic soils, while Gemmatimonadetes was less abundant (Orr et al, 2015). Our results support this findings as Kruskal-Wallis test found Chloroflexi (which represented, on average, 0.018% of the community) to be one of the few phyla that had significantly different abundances in organic and conventional soils, being more abundant in organic soils.

Some long-term studies have found significant differences regarding bacterial diversity of organic vs conventional farming soils. A 21-year study revealed that microbial biomass and activities were enhanced in organic farming soils (Fließbach et al, 2006). Hartmann et al (2006) discovered differences in the genetic structure of bacterial communities in both systems. Van Diepeningen et al (2006) found that the bacterial species richness increased over years of organic management. Esperschütz et al (2007) studied the bacterial response to organic and conventional farming practices by analysing the lipid profiles of bacterial communities in both farming systems, and concluded that there were highly significant differences explained by the soil treatment.

However, even if our study revealed structural differences among the soil samples associated to the two farming systems, the differences were not as visible as expected after treating the soil with such different products, as the farming system accounted for only the 6% of the variance in the beta diversity plots, and low yet significant R values were found in the Adonis test for the “treatment” category.

There are some reasons that might explain why differences between organic and conventional soil microbiomes were not as visible as might be expected. On the one hand, the three sampling campaigns might have been performed in a too short timespan (3 months) for the microbial communities to change significantly. This would explain the lack of significant differences between the communities of the samples taken at different months. In fact, the importance of an extensive spatiotemporal sampling has been highlighted for environmental phylogenetic surveys (Zarraonaindia et al, 2013). For instance, some studies have observed a lack of significant community differences within soil samples collected the same year, while big differences were observed between different years (Zarraonaindia et al, 2015).

In addition, the three sampling campaigns were performed at a very early stage of both the development of the crops (which would explain the lack of differences between the soil communities related to each plant)

and the implementation of organic and conventional treatments (which would explain why bacterial communities in both farming practices did not differ as visibly as other studies have reported). That is to say, more sampling campaigns in a longer timespan might result in bigger differences between organic and conventional communities.

An interesting explanation for not observing structural differences among the portion of bacteria comprising the most abundant taxa of the community could be that the effect of organic and conventional farming in the community might be more functional than phylogenetic. Different taxa can share specific functional attributes just as closely related taxa might have very different physiologies and environmental tolerances (Fierer, Left et al, 2012), and that might lead to different environments having apparently very similar microbial communities, even though specific microbes might play considerably different roles in the community. In fact, it has been shown that functional diversity is not necessarily predictable from the taxonomic or phylogenetic diversity of communities when comparing vegetated soils, whereas other types of soils show strong correlations between taxonomic and functional diversity (Fierer, Left et al, 2012).

5.2- Alternative method for chloroplast interference assessment

The alternative method for evaluating the chloroplast-amplicon content of a sample after the PCR turned out to be accurate and has some advantages in comparison with the method described by Rastogi et al (2010). On the one hand, our alternative method makes chloroplast interference assessment possible without the need of endonuclease *BbvCI*, which makes it faster and cheaper, and requires smaller volumes of PCR product. On the other hand, it can be used for evaluating chloroplast contamination in epiphyte samples of plants whose chloroplast sequence lacks a target for *BbvCI*. Nevertheless, one of the limitations of the alternative method is that it cannot be used in plants whose chloroplast fragment amplified by *27f-YM/Eub518r* primers is not shorter enough than the bacterial amplicons. In addition, the method described by Rastogi et al (2010), despite not been useful in all plant species, could be considered a more direct method than the alternative method we proposed here, as while the former relies on the nucleotide composition of the sequence (direct), the alternative method is based on the size of the amplicons (indirect).

5.3 – Chloroplast interference assessment

The filtration protocol was shown to be the most efficient at excluding chloroplast-DNA before DNA extraction. However, it is not possible at this stage to know, without sequencing the sample, whether this method recovers the whole community, without excluding any particular taxa that might, for instance, live in a biofilm.

6- CONCLUSIONS

- The results of the present study evidenced that there are significant differences in the bacterial community structure of organic and conventional agricultural soils, particularly influencing the abundances of the rare fraction (low abundance taxa) of the community.
- We hypothesize that the farming system might be influencing more the functionality rather than the phylogenetic structure of the community.
- A deeper analysis on the effects of conventional and organic farming systems could be done by combining the results of the present study (soil communities) with the results of the analysis of aerial tissues of tomato and chard that is yet to be conducted.
- Considering that the filtration protocol showed to be the most efficient at excluding chloroplast DNA prior to DNA extraction, I would suggest that it be used in the processing of the tomato and chard tissue samples that were taken in Beotegi.

7- ACKNOWLEDGEMENTS

I am grateful to Dr. Andone Estonba for giving me the chance to undertake such an interesting and inspiring project. Also, this work would not have been possible without the guidance of Dr. Iratxe Zarraonaindia. Her constant support and expert advice, in addition to her thorough revisions have been of invaluable importance. My sincere thanks go to the rest of the lab members, who have always been willing to help with the technical aspects of the project. Particularly, I would like to thank Mikel Aguirre, Iratxe Montes, Naiara Zorrilla, Maialen Sistiaga, David Abad, Anne Olazabal, Jone Michelena and Francisco Llaverro. I would also like to thank Fernando Rendo and Irati Miguel for their helpful suggestions and their careful work in SGIker.

8- BIBLIOGRAPHY

Bulgarelli D, Rott M, Schlaeppi K, van Themaat EVL, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E, Peplies J, Gloeckner FO, Amann R, Eickhorst T, Schulze-Lefert P. *Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota*. Nature (2012) 98:488. Doi: 10.1038/nature11336.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Gonzalez Pena A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. *QIIME allows analysis of high-throughput community sequencing data*. Nature Methods (2010) 7: 335-336.

Esperschütz J, Andreas Gattinger A, Mäder P, Schloter M, Fließbach A. *Response of soil microbial biomass and community structures to conventional and organic farming systems under identical crop Rotations*. FEMS Microbiology Ecology (2007) 61 26–37. doi:10.1111/j.1574-6941.2007.00318.x.

Fierer N, Lauber C, Ramirez K, Zaneveld J, Bradford MA, Knight R. *Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients*. International Society for Microbial Ecology (2012) 6: 1007-1017.

Fierer N, Left JN, Adamds BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso G. *Cross-biome metagenomic analyses of soil microbial communities and their functional attributes*. PNAS (2012) 21390-21395.

Gilbert JA, Jansson JK, Knight R. *The Earth Microbiome project: successes and aspirations*. BMC Biology (2014) 12: 69. Doi: 10.1186/s12915-014-0069-1.

Hanshew AS, Mason CJ, Raffa KF, Currie CR. *Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities*. Journal of microbiological methods (2013) 95(2): 149-155. Doi:10.1016/j.mimet.2013.08.007.

Hartmann M, Fließbach A, Oberholzer HR, Widmer F. *Ranking the magnitude of crop and farmingsystemejects on soil microbial biomass and genetic structure of bacterial communities*. FEMS Microbiology Ecology (2006) 57: 378–388. Doi: 10.1111/j.1574-6941.2006.00132.x.

Kembe SW, Wu M, Eisen JA, Green JL, *Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance*. PLoS Computational Biology (2012) 8(10): e1002743. Doi: 10.1371/journal.pcbi.1002743.

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. *Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies*. Nucleic Acids Research (2012) 1–11. Doi: 10.1093/nar/gks808

Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL. *Practical innovations for high-throughput amplicon sequencing*. Nature Methods (2013) 10: 999–1002. Doi: 10.1038/nmeth.2634

Martins G, Lauga B, Miot-Sertier C, Mercier A, Lonvaud A, Soulas ML, Soulas G, Masneuf-Pomarede I. *Characterization of Epiphytic Bacterial Communities from Grapes, Leaves, Bark and Soil of Grapevine Plants Grown, and Their Relations*. PLoS ONE (2013) 8(8): e73013. Doi:10.1371/journal.pone.0073013.

Orr CH, Stewart CJ, Leifert C, Cooper JM, Cummings SP. *Effect of crop management and sample year on abundance of soil bacterial communities in organic and conventional cropping systems*. Journal of Applied Microbiology (2015) 1364-5072.

Ottesen AR, Gonzalez A, White J R, Pettengill JB, Li C, Allard S, Rideout S, Allard M, Hill T, Evans P, Strain E, Musser S, Kinght R, Brown E. *Baseline survey of the anatomical microbial ecology of an important food plant: Solanum lycopersicum (tomato)*. BMC Microbiology (2013) 13:114.

Rastogi G, Tech JJ, Coaker GL, Leveau JHJ. *A PCR-based toolbox for the culture-independent quantification of total bacterial abundances in plant environments*. Journal of Microbiological Methods (2010) 83: 127-132.

Schloss PD, Handelsman J. *Biotechnological prospects from metagenomics*. Current opinion in Biotechnology (2003) 14: 303-310.

Thomas T, Gilbert J, Meyer F. *Metagenomics - a guide from sampling to data analysis*. Microbial Informatics and Experimentation (2012) 2: 3.

Van Diepeningen AD, de Vos OJ, Korthals GW, van Bruggen AHC. *Effects of organic versus conventional management on chemical and biological parameters in agricultural soils*. Applied Soil Ecology (2006) 31: 120–135.

Wooley JC, Godzik A, Friedberg I. *A primer on Metagenomics*. PLoS Computational Biology (2010) 6(2): e1000667. Doi: 10.1371/journal.pcbi.1000667.



Zarraoinandia I, Smith DP, Gilbert JA. Beyond the *genome: community-level analysis of the microbial world*. *Biology and Philosophy* (2013) 28:261–282. Doi: 10.1007/s10539-012-9357-8.

Zarraonaindia I, Owens SM, Weisenhorn P, West K, Hampton-Marcell J, Lax S, Bokulich NA, Mills DA, Martin G, Taghavi S, van der Lelie D, Gilbert JA. *The Soil Microbiome Influences Grapevine-Associated Microbiota*. *mBio* (2015) 6(2): e02527-14. doi:10.1128/mBio.02527-14.