MASTERARBEIT | MASTER'S THESIS

Titel | Title

Comparing host DNA depletion methods for seagrass leaf microbiome analysis

verfasst von | submitted by

Stefan Eckensperger BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 830
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Molecular Microbiology, Microbial Ecology and Immunobiology
Betreut von | Supervisor:	Assoz. Prof. Dr. Jillian Petersen

Acknowledgments

I first of all want to thank Prof. Jillian Petersen, who welcomed me into her group from the first day, all the way back in February 2022. I am deeply grateful for all the opportunities she provided me with in her group, be it as an intern, a student assistant, or a tutor in her field and lab courses. And finally, of course, for my master's thesis, which she enabled me to do in her group as well. She has always been extremely supportive of me and was available for all kinds of answers and assistance whenever I needed it.

I also want to thank Dr. Hanin Alzubaidy, my lab supervisor. She created this project and allowed me to work on it. She helped me with my presentation for the DOME seminar and my PhD applications. In addition, she let me work on other projects of hers, helping me learn more scientific skills. I want to thank her for taking me on a sampling trip.

I definitely want to thank Philipp Schmelz, my amazing lab partner (or lab husband, as the other master's students called us). We worked together extremely closely during our master theses, and he has always been amazingly helpful. During the period of almost two months I spent in the hospital, he continued with work on both our theses, so without him, I would not be finished even now. But he has also become a good friend outside of academia, with us helping each other survive our studies.

Before I started working with Dr. Alzubaidy, I worked on a different project with Cristina Alcaraz. I want to thank her for supporting me throughout my work, even though nothing came of it in the end. I am very grateful for her help in starting me off at DOME and establishing myself in the lab here.

I also want to thank the TA team at DOME, who were willing to help whenever I had questions about working in the lab. The same is true for Martin Kunert. Even though he is no longer a member of our group, he was always extremely helpful and generous with his time, especially when I struggled with microscopy. I also want to thank all the great people in Prof. Petersen's group for their support and kindness.

For the sequencing, I want to thank the JMF, especially Julia Ramesmayer and Sara Malinowski. They helped a lot with all my laboratory-related problems with DNA extraction and sequencing, answering tons of stupid questions. I also want to thank Dr. Petra Pjevac for organising everything about our sequencing project.

Another group of people I want to thank are all the other master students (and all the other people at the master students' office) at DOME. We have become a great community, great friends, also spending parts of our free time together and helping each other when things sometimes do not work out as they should. The atmosphere in the office might not always have been greatly conducive to work, but it was always a great pleasure to be there. A special thank you goes to Nora Wittmann, who helped a lot with sample processing, despite the fact that she is not even in our group.

Finally, I want to thank everybody in my life who helped me get to this point. First and foremost, my girlfriend, Sophie. She has absolutely no idea about microbiology and genetics, but that does not stop her from being incredibly supportive of everything I am doing. I also want to thank my parents for their support, but also their patience. Now they can finally stop asking when I will be done with my thesis.

Zusammenfassung

Seegräser sind Meerespflanzen. Als Schlüsselarten bilden sie ein Habitat für andere Organismen. Wie bei allen Pflanzen leben auf ihnen auch Mikroorganismen, die eine essentielle Rolle für ihr Überleben spielen. Die Vielfalt des Mikrobioms zu untersuchen ist essentiell, um seine Rolle zu verstehen. Dies wird aber durch die großen Mengen von Host-DNA in Proben erschwert, da diese die mikrobielle DNA übertönen. Es gibt verschiedene Methoden, die Host-DNA reduzieren. Diese Kits wurden aber für menschliche Proben optimiert und nicht auf anderen getestet. Ich habe zwei unterschiedliche Methoden ausprobiert, um die Repräsentation von Prokaryoten zu erhöhen: 1.) kommerziell erhältliche Peptid-Nukleinsäure-PCR-Blocker, und 2.) kommerziell erhältliche Kits, die entwickelt wurden, um Host DNA zu entfernen. Der Erfolg dieser Methoden wurde mittels Sequenzierung des 16S rRNA-Gens überprüft. Dabei wurden signifikante Unterschiede in den erzielten Mikroben-Host-Verhältnissen entdeckt. Es gab auch Unterschiede in der vorhandenen Menge von mitochondrialen und Chloroplasten-Sequenzen, was auf unterschiedliche Effizienz der Methoden bei deren Entfernung hindeutet. Als die beste Methode zur Maximierung des Mikroben-Host-Verhältnisses habe ich den QIAamp Kit identifiziert, der aber auch Grenzen hat. Trotzdem empfehle ich diesen Kit zur Sequenzierung des Metagenoms von Seegrasblättern. Schon aus meinen Ergebnissen war es möglich, Einsicht in das Seegrasblattmikrobiom zu gewinnen. Sie zeigten Organismen, die noch nie auf Seegras gefunden worden waren. Diese Mikroorganismen spielen eine wichtige Rolle für andere Meerespflanzen, unter anderem schützen sie sie vor Krankheiten und helfen bei der Makronährstoffgewinnung. Diese Studie ist die erste Vergleichsstudie über Host-DNA-Entfernung in Pflanzen, sie etabliert ein Protokoll, das auch über Seegras hinaus angewandt werden könnte.

Abstract

Seagrasses are marine plants that are cornerstone species in their ecosystems, creating a habitat for other organisms. Like all plants, they host microbes integral to their ecology. Investigating the diversity of any microbiome is essential to understanding its role. However, this is challenging due to large amounts of host DNA, which obscure prokaryotic DNA. Various methods are available to deplete this host DNA. Most of these methods were optimised for human samples and have not been tested on other samples. I tested two main methods of increasing the representation of prokaryotes: 1.) commercially available peptide nucleic acid PCR blockers, and 2.) commercially available DNA extraction kits designed to remove host DNA. The success of these was assessed by sequencing of 16S rRNA genes, which revealed significant differences in the resulting microbe:host sequence ratios. There were also differences in the representation of mitochondrial vs chloroplast sequences, which might indicate different efficiencies of the kits in removing these. I identified the QIAamp kit as the best method for maximizing the microbe:host ratio in 16S rRNA gene amplicon sequencing, even though it has its limitations. I recommend using this kit for metagenome sequencing of the seagrass leaf microbiome. Already, my results provided insight into the leaf microbiome, revealing microorganisms previously undetected on seagrass leaves. These microorganisms are known to play significant roles in other marine plants, including disease protection and macronutrient acquisition. This study represents the first comparative analysis of host depletion in plants, establishing a protocol that could be applicable beyond seagrasses.

Table of Contents

List of abbreviations

16S	the subunit of rRNA that has a sedimentation coefficient of 16 Svedberg
ASV	amplicon sequencing variant
BLAST	basic local alignment search tool
CAP	constrained analysis of principal coordinates
DNA	desoxy-ribo-nucleic acid
DNase	ribonuclease, DNA-degrading enzymes
JMF	Joint Microbiome Facility of the University and the Medical University of Vienna
mPNA	mitochondrial PNA
NMDS	non-metric multidimensional scaling
PBS	phosphate-buffered saline solution
PCoA	principal coordinates analysis
PCR	polymerase chain reaction
PNA	peptide nucleic acid
pPNA	plastid PNA
RNA	ribo-nucleic acid
rRNA	ribosomal RNA

Figure 1. ***Pictures of a seagrass meadow****. These pictures (© Prof. Jillian Petersen) illustrate a common seagrass meadow. They were taken in a meadow of Zostera noltii at the Plaža Sveta Katarina in Ankaran, Slovenia. The first two panels show the extent and density of the meadow, the others provide examples of the biodiversity in seagrass meadows, showing different species of bivalves, sea anemones and sea urchins, that thrive in this ecosystem. The epiphytes living on the leaves are also visible.*

Introduction

Seagrass

Seagrasses are a paraphyletic group comprising four families in the order Alismatales. They are unique as the only angiosperms, the only flowering plants, in the oceans [1, 2]. Globally, there are approximately 70 species of seagrass [2, 3]. They live in dense meadows close to the coast (Figure 1.).

In the northern hemisphere, *Zostera marina* is the most abundant species of seagrass. It is the dominant seagrass species of the northern Atlantic Ocean [2]. *Zostera noltii* is another important seagrass species, especially in the eastern Atlantic Ocean and the Mediterranean Sea [4, 5], where it often resides at the interface of land and sea due to its capability to survive in intertidal spaces [6]. Like all *Zostera* species, they both have large underground networks of rhizomes that connects large patches of clonal plants and anchors them solidly to the ground [7].

Seagrass ecosystems are important for the environment and humans, but are threatened by climate change

Seagrass meadows form the basis for highly diverse ecosystems [8–10]. They slow down the flow of water around them, capture particles from the water and generally make their habitat more hospitable. This helps many other organisms that might otherwise not be able to thrive, like many species of fish or bivalves [11]. This constructed ecosystem enables the survival of microorganisms that benefit the seagrass, for example many different species of nitrogen fixing bacteria living in the sediment. They interact with the seagrass roots and serve similar roles as their counterparts do for terrestrial plants [12, 13].

Seagrasses also provide ecosystem services to humans: Seagrass meadows serve as nursing areas for various fish species, many of them of economic value [2, 8]. They also stabilize coastal sediments [2, 8], and filter contaminants [14] and pathogens [15–17] from the water. In total, seagrass meadows provide 7,500 $ of value per hectare per year [18].

The seagrass ecosystem is under strong pressure all around the globe. They are threatened by eutrophication, rising sea levels, and rising ocean temperatures [10]. This leads to about 15% of all seagrass species being endangered, including *Zostera marina* [19, 20]. Because of seagrasses'

roles as keystone species, that means many other species are threatened indirectly [19]. Because many fish species inhabit seagrass meadows during some part of their life cycle, the potential effects reach far beyond the seagrass meadows themselves [19].

Seagrass harbours a diverse microbiome serving many roles

Seagrass functions as a habitat for its own bacterial community [15]. It provides a large amount of organic carbon to its surrounding water [21] and sediment [22]. This makes the surroundings a nutrient-rich habitat for bacteria [23]. Seagrass also enriches nitrogen fixing bacteria in the soil around its roots [24], which in turn provide nitrogen to the seagrass [25, 26]. Other members of the seagrass microbiome detoxify the surrounding sediments [27–29]. The seagrass microbiome also plays a large role in the health of the seagrass plants [30]. It protects the plant against pathogens [31–34] and produces antimicrobial compounds [35]. It also inhibits the growth of algae on the plant surface [36, 37].

The seagrass microbiome is structured into three distinct sections: the phyllosphere (the leaf surface), the rhizosphere (the root surface and its immediate surroundings), and the endosphere (the inside of the plant), the last of which is again different in the plant parts above and below the seafloor [15]. The phyllosphere is very similar in composition to the community of the surrounding water. It has been suggested that it forms by recruitment and selective enrichment from the surrounding environment [15, 38–40]. The leaf microbiome is less diverse than the surrounding seawater and consistently shows specific taxa enriched compared to its surroundings, irrespective of sampling site [40, 41].

Host depletion

What is host ‘contamination’?

To analyse the microbial community of any given habitat without first having to cultivate the microbes living there, possible approaches include high-throughput sequencing of PCR-amplified marker genes, most commonly the 16S rRNA gene, or shotgun metagenome sequencing, which means sequencing all DNA contained in an environmental sample, with the aim of subsequently assembling genomes of the organisms contained in the sample computationally. Both approaches have been used to great success in many different environments, including seagrass meadows [42–47].

In metagenomes, especially those from eukaryotic hosts, analysing the associated prokaryotic communities is difficult. The genome of the host is usually larger than any prokaryotic symbiont genome, therefore, just a few host cells can lead to the host DNA ‘drowning out’ the prokaryotic DNA [48–51]. In human samples, this can lead to anything between 10 and 90% of all sequencing reads being assigned to the human genome [48]. This is a problem because it can lead to rarer organisms becoming invisible, and because it wastes sequencing power on sequencing the host genome, which is already available in many cases.

Even in 16S rRNA sequencing, where eukaryotic reads are not sequenced, host ‘contamination’ is still present. This is because the host contains 16S rRNA genes from its mitochondria, and – in plants – chloroplasts [51]. This can lead to those host organelles making up more than 90% of all reads recovered from plants [51–53].

There are different approaches for host depletion

To solve this problem, there is no easy, catch-all solution. There are different methods to try to solve it, but which one is best depends on many factors. Every sample is different, and for every specific application on every specific sample, the best-suited method can be different. This means that for every sample, different methods have to be tested and optimised before picking one. The most important factor determining which approach to use is whether amplicon sequencing or full metagenomics are to be done.

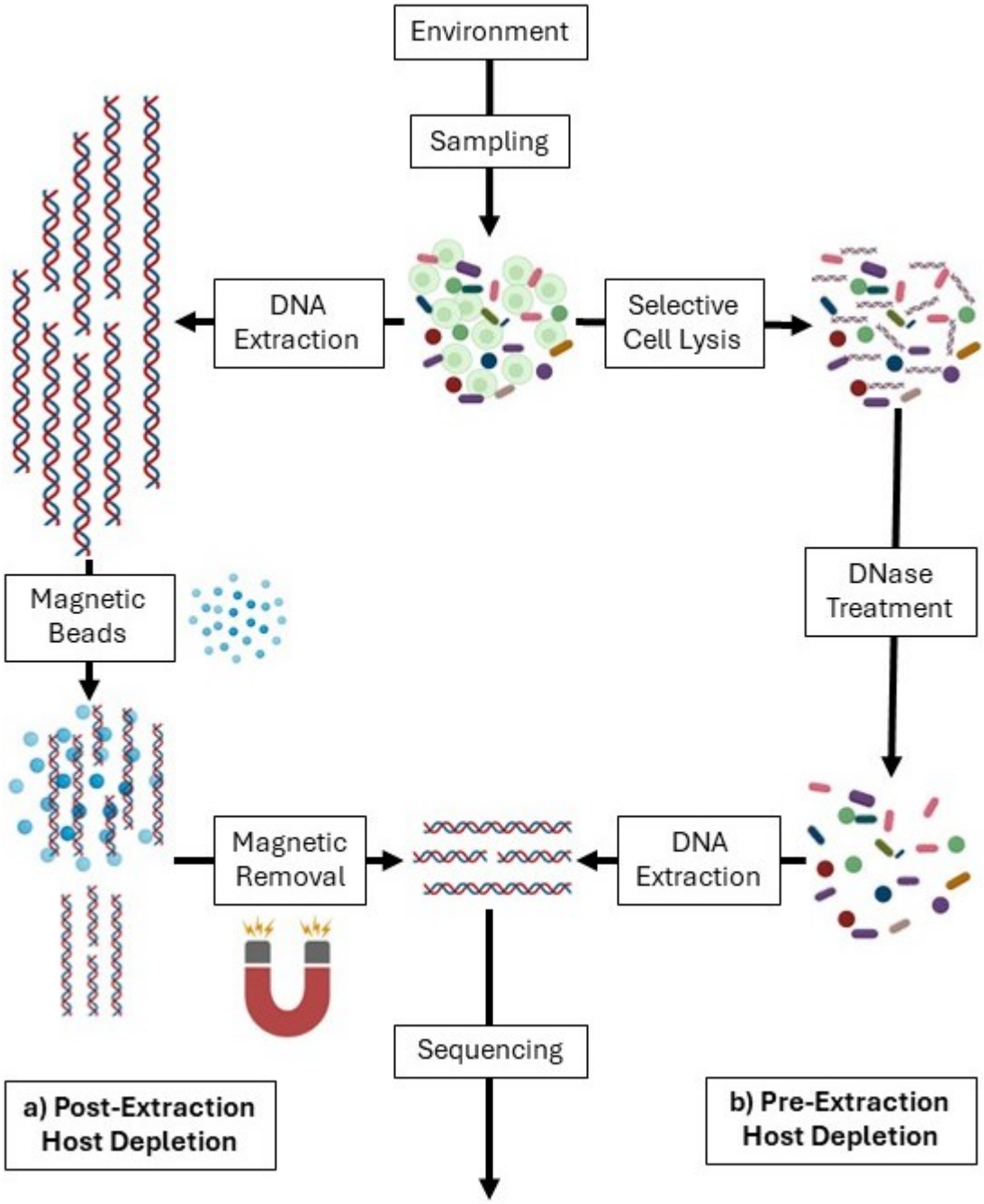


Figure 2. ***Overview of the two main approaches of host depletion.*** *Overview of the general workflow for the two main categories of host depletion methods to go from the environmental sample to the DNA for sequencing. a) on the left shows post-extraction host depletion, where DNA is extracted before eukaryotic methylated DNA (for example, it can be any pattern specific to eukaryotic DNA) is removed by using magnetic beads that bind these methylation sites. b) on the right shows pre-extraction host depletion, where first all eukaryotic cells are lysed selectively, then the now-free DNA in the sample is lysed by DNase treatment, and afterwards, standard DNA extraction is performed on the remaining prokaryotic cells.*

For amplicon sequencing, the method most easy to use is sequencing with PNAs, peptide nucleic acid clamps. It is based on depleting specific, known sequences. This approach does not affect the extracted DNA. It is isolated the usual way and prepared for sequencing. Before sequencing, PNAs are added. PNAs are synthetic compounds linked to a specific sequence of nucleotides [54, 55]. Those nucleotides can bind DNA, blocking specific sites. This pairing happens selectively, down to the level of single nucleotide changes [56]. PNAs were used first as PCR blockers because of their selectivity [55, 57], but are also used to inhibit sequencing of specific sequences. If one PNA specific to the 16S rRNA gene in the mitochondria of the host (called mPNA), and, for plants, another one specific to the 16S rRNA gene in the plastids of this host (pPNA), are added to the DNA, they bind the relevant sites, the sequencing reaction cannot happen, and they do not show up in the final results [56]. I will refer to this as host blocking in this study. For metagenomics, PNAs or similar methods are not applicable.

For host depletion in metagenomics, various kits aimed at maximizing the proportion of bacterial and archaeal DNA in the sample and minimizing eukaryotic DNA are commercially available. Those kits give wildly different results when tested on different samples [48–50, 58–63]. They use different properties that differentiate prokaryotic and eukaryotic cells and fall into two categories. The first category, post-extraction host depletion kits, use the different methylation patterns of prokaryotic and eukaryotic DNA to separate them after DNA extraction. They use beads or columns that selectively bind either methylated or non-methylated DNA (Figure 2.). Those kits are normally very quick to use, and are able to process very low sample volumes, but require high-quality, large-fragment DNA to work, which is often not possible to obtain from environmental samples. The most widely used kit from this category is the NEB Next Microbiome DNA Enrichment Kit by New England Biolabs [61]. The other category, pre-extraction host depletion, uses a different approach. It uses enzymes and mechanical lysis to selectively lyse eukaryotic cells in the sample before DNA extraction. This mixture of prokaryotic cells and free DNA is treated with DNases to degrade all free DNA, leaving behind just the prokaryotic cells. The DNA from these cells is then extracted (Figure 2.). This category of kits, as opposed to the other, tends to be very time-consuming, and requires sizeable amounts of biomass. On the other hand, the kits are able to extract DNA from any kind of sample, which is especially useful for environmental samples. This category includes most of the commercially available kits, for example the QIAamp DNA Microbiome Kit by Qiagen, and the HostZERO Microbial DNA Kit by Zymo Research [61].

For both categories of kits, but especially for the pre-extraction kits, there is the additional risk that they might introduce biases into the observed community composition. There is a risk that the selection methods affect specific clades of prokaryotes more than others, leading to them being more or less represented in the dataset than in the environment [50, 61]. All these kits have to be optimised and tested for every new sample type that is to be used. This is especially true for all samples that are not human in origin, as this is what the kits were all optimised for. Currently, though, there is no data available that compares host depletion in seagrass samples. In fact, there is no study at all that uses even a single host depletion method on seagrass samples. Furthermore, there are hardly any studies that use host depletion methods on plant material, neither comparative studies nor studies working with even a single host depletion method.

Aims of the study

Seagrass ecosystems are very important to marine ecology. The microbiome of seagrass plays an essential role in its survival. It is still very much understudied, especially the leaf microbiome. According to the current state of knowledge, the leaf microbiome forms mostly by selective assembly from the surrounding water [30, 40].

In DNA extracted from environmental samples, eukaryotic DNA often drowns out prokaryotic DNA, especially when sampling the microbiome of a eukaryotic host. This problem is referred to as host 'contamination'. There are many different approaches to solve this problem, but to the best of my knowledge, none of them have ever been tried on seagrass samples.

In this study, I aimed to compare different protocols for the reduction of host 'contamination' in seagrass leaf samples for both 16S rRNA gene amplicon sequencing and metagenomics. I wanted to identify the protocol that can deplete host tissue the best, while maintaining the prokaryotic community composition. To do that in a time-efficient way, I used 16S rRNA gene amplicon sequencing to quickly analyse the community. As a proxy for host depletion, I used the relative abundance of host organelles in the sequenced samples.

I wanted to establish host depletion as an approach to look deeper into the microbiome of the seagrass leaves. Recently, more data has emerged, showing that the seagrass leaf microbiome is made up of a consistent, selected core microbiome. To be better able to study the potential interplay between that core microbiome and the seagrass, deeper sequencing will be necessary. On the leaves, where microbial abundances are quite low and thus easily drowned out by host 'contamination', host depletion is even more important for their study.

Materials and methods

Sampling site and sample collection

Zostera marina samples were collected on October 12th, 2023, at one site in the Ria de Ferrol at Ferrol, Spain, just off the Playa del Castillo (43.27.51.0 N, 8.17.01.5 W). The seagrass meadow at this site is in a subtidal zone, approximately 5 m below the sea-level. The plants sampled by the divers were spread out across a larger area: A perimeter of 30 m by 20 m was marked on the seafloor. Inside this perimeter, beginning in one corner, three smaller perimeters were marked. Each of these perimeters was 3 m by 2 m, oriented in the same way as the large one. The three smaller perimeters were each 3 m apart from each other. From each of the three small perimeters, 5 individual seagrass plants were taken, being careful to take not only the leaves, but to also get the rhizome with as much of the roots as possible, and not to damage both the plant and the surrounding environment, choosing those that were somewhat close to each other. In addition, from outside the three small perimeters, another ten plants were taken, selecting plants that were far apart from each other. The sampling scheme is illustrated in Figure 3. Each plant was collected in a separate plastic bag and taken to the lab for further processing.

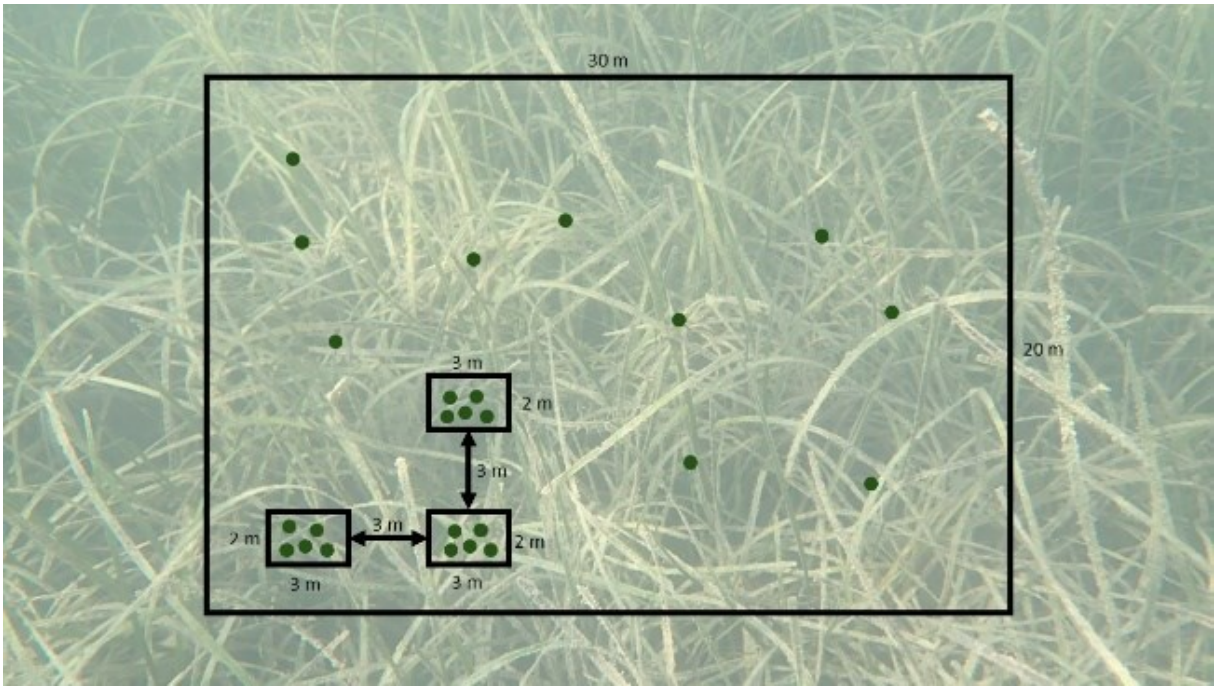


Figure 3. ***Sampling schemes used for seagrass sampling.*** *Illustration of the sampling schemes used for sampling of* Zostera marina *seagrass in Ferrol, Spain. Each dot represents one individual plant taken as a sample.*

Sample processing

Samples were processed by me and the rest of the team present during field work, immediately after sampling in the field. For the leaf samples, we removed the very outermost leaves that were very short or already damaged. From the remaining leaves, we took the outermost one. This leaves we rolled up, placed in a 2 ml Eppendorf tube and snap-froze the tube in liquid nitrogen. We stored the tubes at -75°C for the transport and until further processing. During this processing, we took care to avoid any kind of cross contamination between samples. We washed all surfaces and tools with distilled water and disinfected them with 70% ethanol after every single individual plant. We also replaced the used scalpel blades and gloves after each plant.

Back in our home laboratory, we processed the *Zostera marina* samples further by grinding them in liquid nitrogen. We filled an autoclaved ceramic mortar with liquid nitrogen, added the sample with a pair of tweezers, and ground it into a fine powder. The powder resulting from every individual plant, we then distributed equally to three 2 ml Eppendorf tubes, each of which we weighed. We called them aliquots A, B, and C. Then, we stored those samples again at 75°C until further analysis. During the grinding, we again took extreme care to avoid any kind of cross contamination. We disinfected all surfaces with 70% ethanol and cleaned them with RNase away between each sample. We replaced and disinfected our gloves after every sample. We washed all equipment (mortars, pistils, spatulas, tweezers) with water and soap. After washing, we wrapped the equipment and autoclaved it.

DNA extraction

I performed DNA extraction on one aliquot of every ground-up plant (aliquot A). I used the DNeasy PowerSoil Pro Kit from Qiagen. The extraction was mostly performed according to the manufacturer's instructions, with the following exceptions: I took the full weight of the sample aliquot in the initial step, not only 250 mg. After adding the solution CD2, I left the samples to incubate for 5 min, the same after adding solution CD3. I performed the elution in 50 µl of solution CD6, and, after adding the buffer to the membrane, left the membrane to incubate for 5 min. After this elution step, I pipetted the eluate again onto the membrane, left it to incubate for 5 min, and eluted again. I split the DNA in two aliquots (A1 and A2). The full, detailed protocol I used is found in appendix 1.

I evaluated the quality of the extracted DNA with different methods: For evaluating the DNA concentration, I used both a Nanodrop 1000 and a Qubit 4, both from ThermoFisher Scientific. The DNA fragment size I analysed with a 4150 TapeStation System from Agilent.

Host blocking

For every plant, to one part of the DNA extracted with the PowerSoil kit (aliquot A2), I added two PCR inhibitors, mPNAs for mitochondria, and pPNAs for chloroplasts, both from PNA BIO. The sequence of mPNA is ggcaagtgttcttcgga, the sequence of pPNA ggctcaaccctggacag.

Host depletion

I performed host depletion with two different host depletion kits. Both of them are pre-extraction host depletion kits. The first kit I used was the QIAamp DNA Microbiome Kit from Qiagen. The other kit was the HostZERO Microbial DNA Kit from Zymo Research.

QIAamp DNA Microbiome Kit

I performed Host Depletion with the QIAamp DNA Microbiome Kit on one aliquot per plant (aliquot B). I followed the manufacturer's instructions, with the following exceptions: To get the sample as a liquid, as was required by this kit, I resuspended each sample in 1 ml of 1x PBS and filtered it through a 40 µm cell strainer afterwards. In exchange, I removed the step containing the end-over-end rotation. I did the bead beating on a FastPrep 24 at 6.5 m/s, 3 times for 1 min each, with the samples kept on ice in-between. After the elution step, I again pipetted the eluate onto the membrane, left it there to incubate again for 5 min and eluted it again. The full, detailed protocol I used is found in appendix 1.

HostZERO Microbial DNA Kit

I performed Host Depletion with the HostZERO Microbial DNA Kit on one aliquot per plant (aliquot C). I performed Host Depletion with the HostZERO Microbial DNA Kit mostly following the manufacturer's instructions, with the following exceptions: To get a liquid sample, as was required by this kit, I resuspended the ground-up sample in 1 ml of 1x PBS and filtered it through a 40 µm cell strainer afterwards. In exchange, I removed the step containing the end-over-end rotation with buffer AHL. I also performed the optional heating step after lysis, as proposed by the manufacturer. After the elution step, I again pipetted the eluate onto the membrane, left it there to incubate for 5 min and eluted it again. The full, detailed protocol I used is found in appendix 1.

Sequencing

I sequenced the following samples: The DNA that had been extracted with the PowerSoil kit (aliquot A1), the DNA that had been extracted with the PowerSoil kit and to which PNAs had been added (aliquot A2), the DNA that had been depleted and extracted with the QIAamp kit (aliquot B), and the DNA that had been depleted and extracted with the HostZero kit (aliquot C). 16S rRNA gene amplicon sequencing was performed by the Joint Microbiome Facility, JMF, of the Medical University of Vienna and the University of Vienna. The ID of the run was JMF-2404-07. The sequencing method chosen was Illumina MiSeq, using the primers 515F Parada and 0806R Apprill. The data was delivered by the JMF with taxonomic assignment already done.

Data analysis

I performed all data analysis in R, version 4.3.2. I investigated sequencing quality using the packages phyloseq, version 1.46.0 [64], and vegan, version 2.6-4 [65]. I handled the 16S rRNA gene amplicon sequencing data with phyloseq [64], microbiomeutilities, version 1.00.17 [66], and microViz, version 0.12.1 [67]. To test host blocking and host depletion for significance, I did a two-sample, two-sided t-test with BSDA, version 1.2.2 [68].

I analysed and plotted α-diversity with phyloseq [64]. The NMDS analysis, the PCoA analysis, and the CAP analysis I performed and plotted with phyloseq [64] and analysed statistically with a permutational analysis of variation with usedist, version 0.4.0 [69], and phytomosaic/ecole, version 0.9-2021 [70]. For the CAP analysis, I additionally used microbiome, version 1.24.0 [71]. The venn diagram analysis was performed with ggvenn, version 0.1.10 [72]. I performed the differential abundance analysis with DESeq2, version 1.42.1 [73].

Most figures were plotted with ggplot2 [74], but I also used gridExtra, version 2.3 [75], and patchwork, version 1.2.0 [76]. To introduce line breaks in figure legends and labels, I used scales, version 1.3.0 [77]. For general data management, I used data.table, version 1.15.0 [78] [79], gt, version 0.10.1 [80], Hmisc, version 5.1-3 [81], and reshape2, version 1.4.4 [82]. Additional packages that are required for use of the main packages used are dplyr, version 1.1.4 [79], knitr, version 1.45 [83], plyr, version 1.8.9 [84], and tidyverse, version 2.0.0 [85]. The full code of the analysis is found in appendix 2.

Results

In this thesis, I wanted to compare different methods to reduce host 'contamination' in seagrass leaf samples. For this, I compared different approaches. The first approach I tried was host blocking with PNAs during sequencing. I then tried two kit-based host-depletion methods. I decided to use the QIAamp and the HostZero kits. I did not use any post-extraction kits, as the most widely used one did not work at all on these samples in preliminary experiments. As a control, I used samples where I performed classic DNA extraction with the PowerSoil kit. This kit is the most widely used kit in the field. In preliminary testing, it worked very well for DNA extraction in our samples.

Host organelles dominate the sequenced seagrass leaf microbiome

I first wanted to see how the prokaryotic community of the seagrass was structured in my sequencing results. For this, I used samples that were not treated with any method to reduce host 'contamination'. The DNA in these samples had only been extracted with the PowerSoil kit and then sequenced. I looked at the average community composition of these samples to determine which families make up the seagrass leaf prokaryotic community.

The leaf microbiome was dominated by host organelles. Almost two thirds of all sequenced reads from the microbiome were identified as coming from host organelles. Almost 50% of reads were coming from mitochondria alone. Chloroplast reads made up about 20% of all reads sequenced. This was about the same amount of reads as were assigned to the 10 most abundant prokaryotic families in the samples.

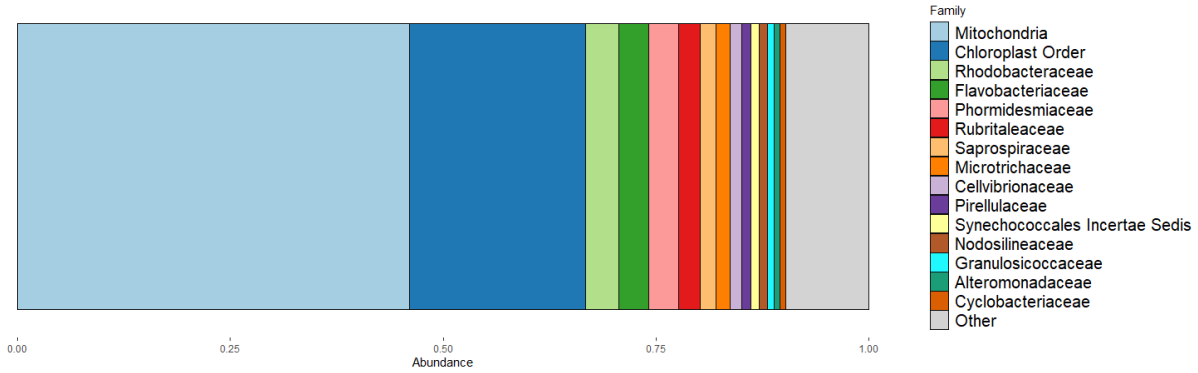


Figure 4. ***Community composition of seagrass leaves.*** *The x-axis depicts the relative abundance in the sample. This figure depicts the community composition, with different colours denoting the 15 most abundant families in the samples. Light blue denotes mitochondria, dark blue chloroplasts.*

The prokaryotic community is dominated by a few families. The most abundant families in the samples were Rhodobacteraceae, Flavobacteriaceae, Phormidesmiaceae, and Rubritaleaceae. Reads from each of them made up between 3 and 5% of all reads sequenced. Together, they made up about one third of all prokaryotic reads sequenced.

These results showed that some approach to reduce host 'contamination' was definitely necessary. If only one third of all sequencing power is used on the prokaryotic community, much of its diversity will likely be missed. The host organelles should be removed from the sample before sequencing. The easiest approach to do that is to use PNAs.

Host blocking with PNAs did not work

First of all, I wanted to determine if host blocking in seagrass was successful. To do that, I calculated the relative abundance of host organelles, mitochondria and chloroplasts, in both treatments. I evaluated the changes in host content between the control and the samples sequenced with addition of PNAs. For easier analysis, I combined all samples that had been treated with the same method.

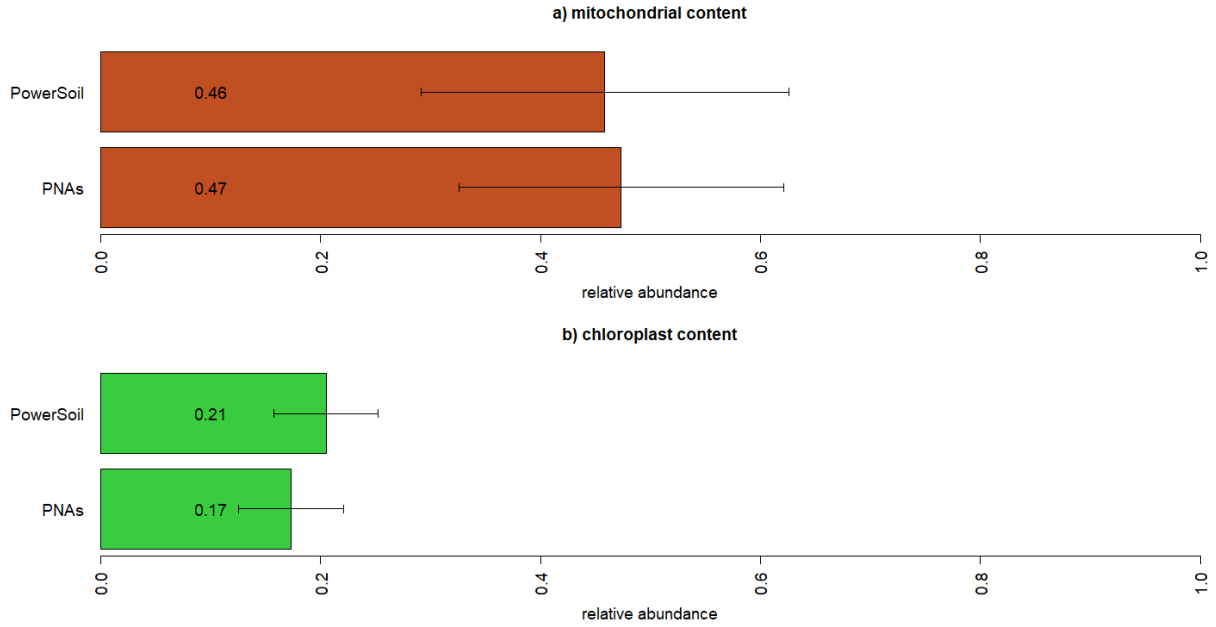


Figure 5. ***Host content of samples sequenced with PNAs.*** *The x-axis depicts the relative abundance in the different samples sequenced with and without PNAs. The y-axis shows the treatment of the DNA, either DNA extraction or host blocking. The brown bars in the upper panel denote mitochondrial relative abundance, the green bars in the lower panel chloroplast relative abundance. The numbers in each bar signify the exact value for that sample. The error bars denote standard deviation. The mitochondrial content is not significantly different between treatments (p=0.7657), while the chloroplast content is decreasing significantly (p=0.03731) when samples are sequenced with PNAs.*

There was no strong effect of PNA addition on the relative host content compared with samples extracted with the PowerSoil kit and sequenced without PNAs (Figure 4.). Mitochondrial content did not change significantly. Chloroplast content did change slightly, but statistically significantly. It decreased from 21 to 17% of all sequences (Figure 4.). Total host depletion caused by the PCR blockers was 4.5% when compared to the host content of the samples sequenced without PNAs. Mitochondria were not depleted at all, their abundance increased by 2.2%. Chloroplasts were depleted and decreased by 19%.

The results show that PNAs are barely able to perform host blocking in seagrass leaves. They were not at all able to block sequencing of mitochondria. While they were able to block sequencing of chloroplasts, this only led to a reduction of chloroplast read by less than a fifth, still leaving most chloroplast reads to be sequenced. This is not enough to efficiently remove host ‘contamination’ from the sample

Host depletion with kits worked well

The QIAamp kit reduces host 'contamination' more efficiently than the HostZero kit

To determine the success of host depletion in seagrass leaves, I evaluated the changes in community composition and host content between the control and the samples depleted with the two host depletion kits. For easier analysis, I combined the samples depleted with one kit to analyse them as one treatment.

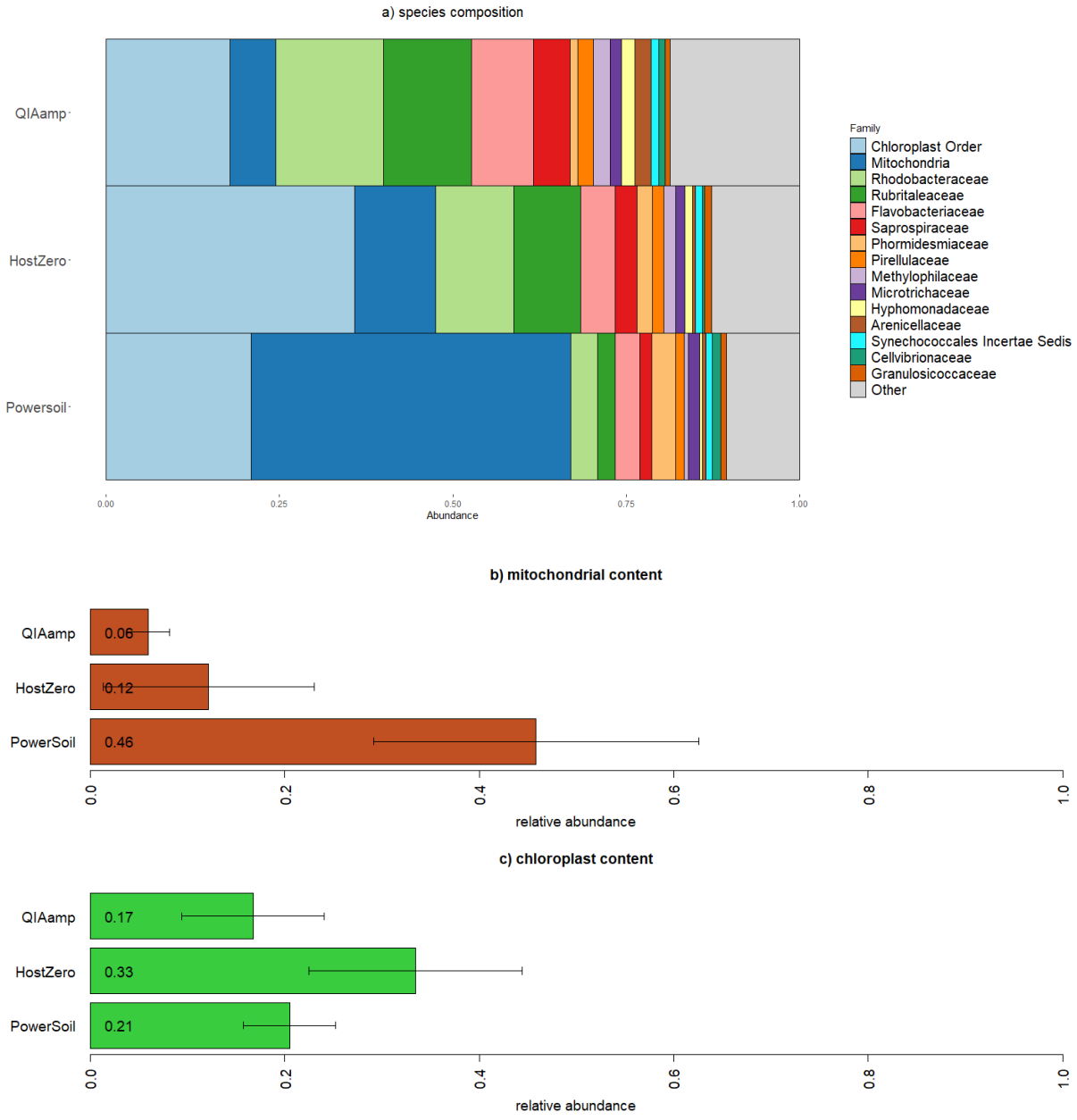


Figure 6. ***Community composition and host content of samples treated with different host depletion kits.*** *The x-axis depicts the relative abundance in the different samples sequenced with and without PNAs. The y-axis lists the method that was used for host depletion. a) depicts the community composition, with different colours denoting the 15 most abundant families in the samples. Light blue denotes mitochondria, dark blue chloroplasts. b) depicts the mitochondrial content of samples treated with each method. The bar denotes mitochondrial relative abundance. The numbers in each bar signify the exact value for that sample. The relative mitochondrial content is significantly decreasing with both HostZero ($p=4.547x10^{-8}$) and QIAamp ($p=3.245x10^{-9}$). c) depicts the chloroplast content of samples treated with each method. The bar denotes chloroplast relative abundance. The numbers in each bar signify the exact value for that sample. The relative chloroplast content is not decreasing significantly with QIAamp ($p=0.1853$) and is even increasing significantly with HostZero ($p=2.985x10^{-4}$).*

Community composition did not vary much between the different treatments. Host depletion decreased the relative abundance of host organelles, while at the same time increasing the relative abundance of most of the other families in the samples. In all host depletion treatments, mitochondria and chloroplasts remained the most abundant representatives.

Even if mitochondria and chloroplasts were the most abundant representatives in all samples, the differences between the samples extracted with the PowerSoil kit, and those treated with host depletion kits was obvious: Of the samples extracted with PowerSoil, mitochondria made up 46% of reads, and chloroplasts made up 21% of reads. The total host content was 67% (Figure 5.). For the two host depletion kits, those figures were much lower.

Samples treated with host depletion kits had a lower host content than samples where DNA was extracted normally. For samples treated with the HostZero kit, total host content was 45% of reads, 12% were mitochondrial, 33% chloroplast reads. For samples depleted with the QIAamp kit, total host content was just 23%, 6% mitochondrial reads, and 17% chloroplast reads. Especially for mitochondria, that was a clear decrease (Figure 5.).

Mitochondria were significantly depleted by both host depletion methods. The HostZero kit depleted mitochondrial reads by 73.9% compared to the samples extracted with the PowerSoil kit, the control ($p=4.547x10^{-8}$). The QIAamp kit decreased the relative abundance of mitochondrial reads even stronger, depleting them by 87.0% compared to the control ($p=3.245x10^{-9}$, Figure 5.).

For chloroplasts, the picture was not as clear. The HostZero kit did not deplete chloroplast sequences, in fact, they were significantly increasing in relative abundance during treatment with this kit, by 57.1% compared to the control ($p=2.985x10^{-4}$). The QIAamp kit did not increase them, but it also did not decrease them significantly. (Figure 5.)

Both host depletion kits were able to deplete host tissue. Both host depletion kits tested significantly decreased the abundance of host mitochondria in the samples, QIAamp more so than HostZero. Neither of them was able to deplete relative chloroplast abundance. Again, QIAamp performed better than HostZero, because during treatment with it, the relative abundance of chloroplast sequences did not increase in the samples.

Host depletion increases the number of prokaryotic groups detected

The goal of host depletion is to enable deeper sequencing of the prokaryotic community without using more sequencing power. This is achieved by avoiding 'wasting' sequencing power to generate reads of the host genome. Therefore, samples sequenced after host depletion should not only show lower host content, but also higher alpha diversity. To analyse this, I calculated different indices of α-diversity and choose three of them: Observed diversity, also called richness, is just the number of species in the sample. Chao1 uses species of which only a single read is present to calculate the actual richness. Shannon, the third index, also takes the evenness of the community into account, not just the diversity.

The α-diversity was similar in all treatments, except for the samples where host depletion was performed with the QIAamp kit. In these samples, all indices of α-diversity were significantly higher (Figure 6.). The median observed diversity for the samples extracted with PowerSoil was around 750, while it was around 825 for the samples depleted with the HostZero kit. For the samples depleted with the QIAamp kit, the single-lowest value was 975, while the median was 1100 ASVs, which was significantly higher than the others ($p=1.932x10^{-5}$). A very similar pattern was observed with the Chao1

index. The difference here was that the median for the samples depleted with HostZero was the lowest and that the difference between the samples treated with QIAamp and the others was less pronounced, but still significant ($p=4.294x10^{-3}$). The same pattern held true for the Shannon index. The median Shannon index for the samples depleted with the QIAamp kit was the highest, being significantly different from all the other (p=8.75x10-7), while it was lower for all the others, but slightly higher for the samples depleted with the HostZero kit (Figure 6.).

The diversity analysis showed that of the two host depletion kits, only one was able to detect significantly more of the diversity of the sample. The QIAamp kit increased all indices of α-diversity significantly. This means that it is not only the more successful kit in regard to host depletion, but also has the intended effect of increasing diversity in the sequencing results.

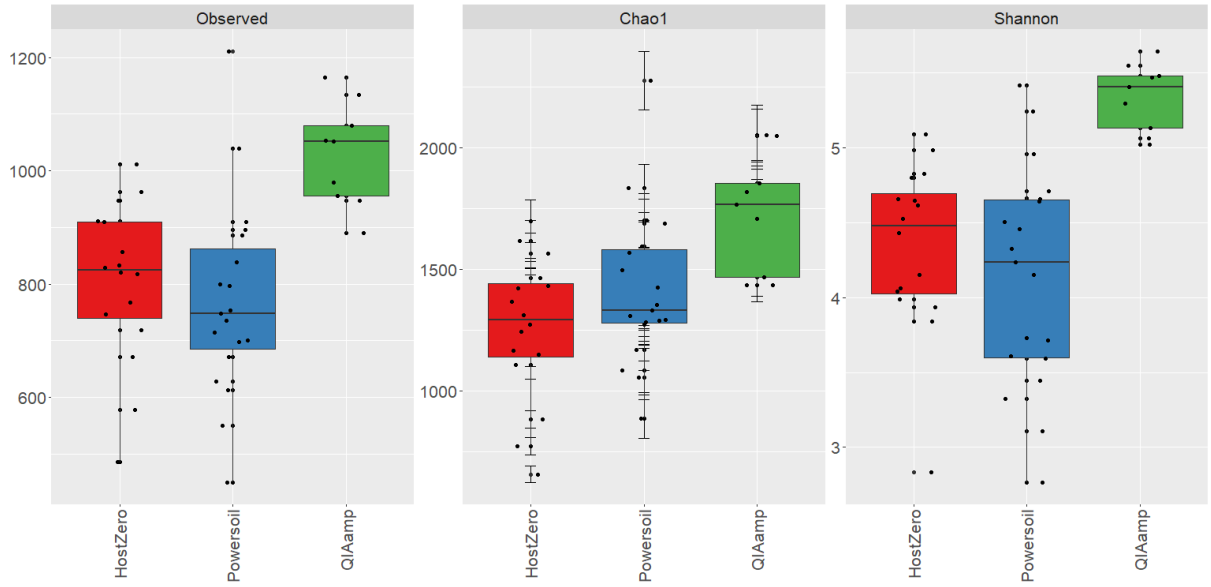


Figure 7. ***α-diversity of samples treated with different host depletion kits.*** *The y-axis depicts the value of the α-diversity index shown in the title of each sub-panel (on the left observed diversity, in the middle Chao1, and on the right Shannon diversity). The x-axis lists the method that was used for host depletion. The boxplots cover the area from the 25th to the 75th percentile, while the arms cover the rest of the samples, excluding outliers. Each point represents one single sample The observed diversity is not increasing significantly for HostZero (p=0.5754), but it is increasing significantly for QIAamp ($p=1.932x10^{-5}$). The Chao1 index is not increasing significantly for HostZero (p=0.1901), but it is increasing significantly for QIAamp ($p=4.294x10^{-3}$). The same pattern also holds true for the Shannon index: it is not increasing significantly for HostZero p=(0.3543), but it is for QIAamp ($p=8.75x10^{-7}$).*

Host depletion introduces biases into the sequencing results

There were large differences in which taxa were detected by which kit

To further investigate the effects of host depletion, I performed a Venn diagram analysis of the taxa identified from the 16S rRNA gene amplicon sequencing. I wanted to see whether there are taxa that are only detected in samples extracted with one or two of the kits, but not the others. I wanted to make especially sure that the two host depletion kits detect most, or all of the organisms detected by normal DNA extraction.

In samples extracted with different kits, there were large differences in the genera detected in the microbiome of the seagrass leaves. Only 446 ASVs at the genus level were detected in samples from all three treatments, out of 765 ASVs in total. That only accounted for 58.3% of all genera in this analysis. There were 64 ASVs, accounting for 8.4% of all ASVs that were detected in samples extracted with both host depletion kits, but not in the samples derived from normal DNA extraction, and 14.2% that were only detected in samples extracted with one of the two host depletion kits.

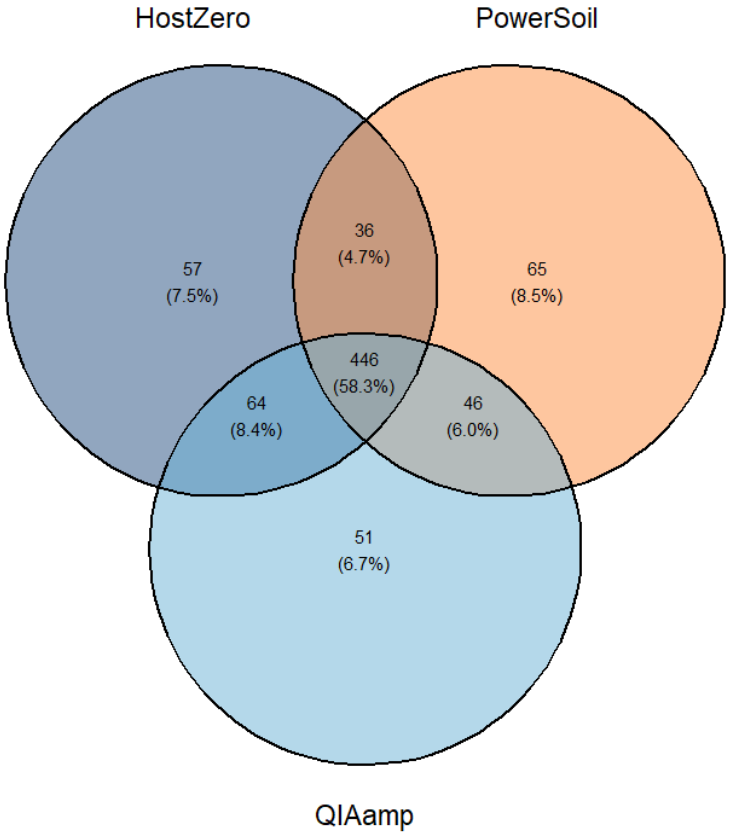


Figure 8. ***Venn diagram of genera present in samples treated with different host depletion kits.*** *The diagram depicts the number of genera that are present in each section of the samples.*

Many genera are not detected in host-depleted samples. 65 ASVs (8.5%) were observed only in the samples that were extracted with PowerSoil. For every host depletion kit, there is a large number of taxa that are missed: The samples only detected by the PowerSoil kit, the samples only detected by the other host depletion kit, and the samples detected by both of these methods. The HostZero kit misses 21.2% of all ASVs, while the QIAamp kit misses slightly less, at 20.7% of all ASVs (Figure 8).

This analysis showed that both host depletion kits miss a sizeable portion of the diversity present in the sample. Even as diversity is increasing for the QIAamp kit, it still does not detect the whole community. In fact, it misses parts of the community that would have been detected with a normal

DNA extraction protocol. This also suggested implications for how well the community structure is detected by host depletion.

Community differences induced by host depletion are only partially due to the removal of host organelles

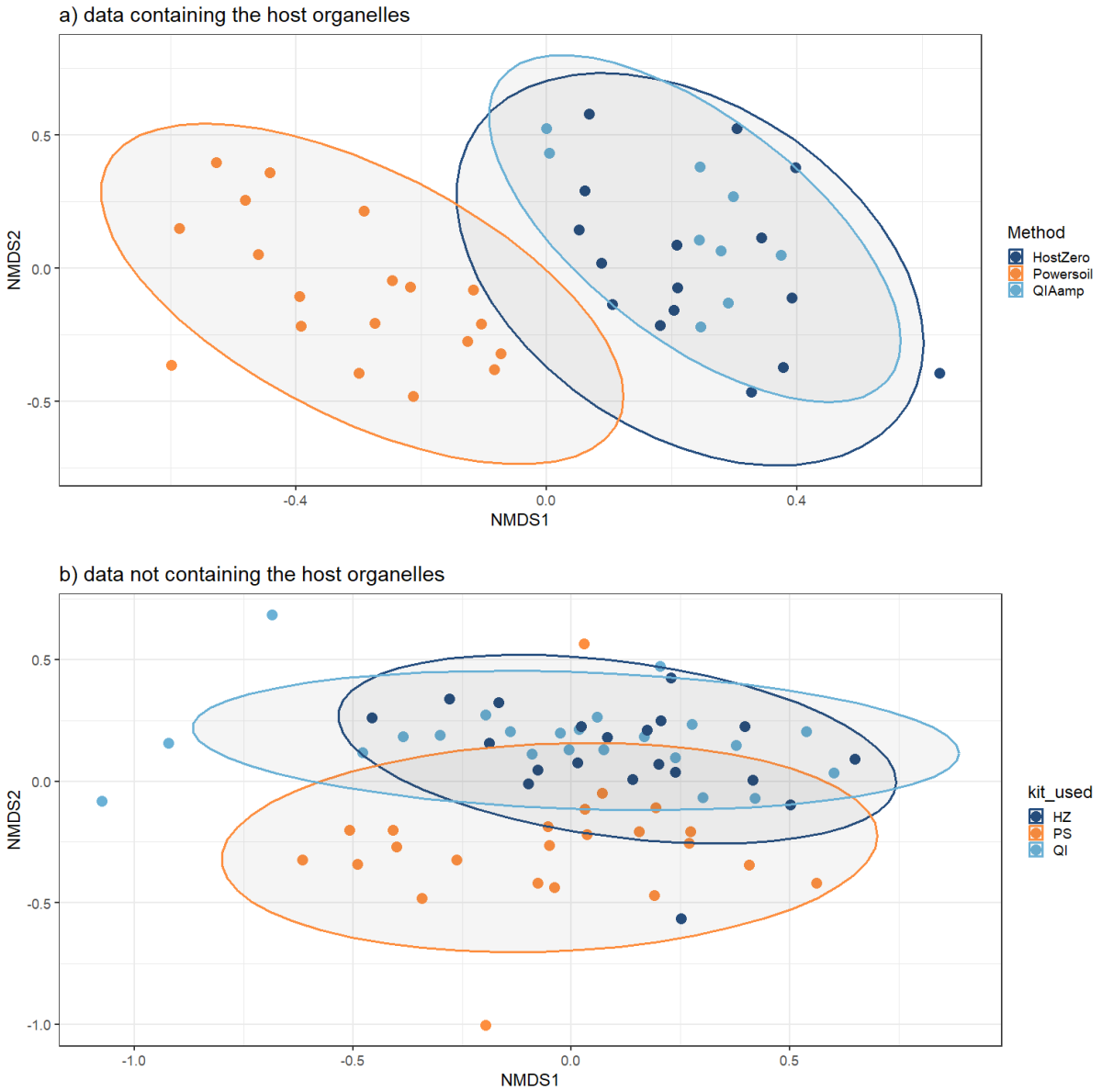


Figure 9. ***NMDS analysis of community composition of samples treated with different host depletion kits.*** *The two axes depict the two dimensions of the NMDS analysis. Each point represents a single sample, with their distances in the plot representing the real distance of their community composition. Different colours represent different treatments. The ellipses represent the 95% confidence interval of each treatment. a) depicts the differences in community composition as they are. All samples are changing significantly with host depletion. The change is clearly significant for HostZero ($p>10^{-4}$, $R^2=0.431$) and QIAamp ($p>10^{-4}$, $R^2=0.403$). b) depicts the differences in community composition as they are, but with the host organelles removed* in-silico. *HostZero ($p=3x10^{-4}$, $R^2=0.158$) and QIAamp ($p=3x10^{-4}$, $R^2=0.129$) are changing the community significantly.*

I wanted to investigate if the community composition was affected by the host depletion kits. To see if there was a change in the community composition, I performed NMDS analysis again to compare the samples treated with different host depletion kits. I constructed two such NMDS analyses. The first was with the sequencing results. The second was with the reads assigned to host organelles removed *in-silico*, to investigate the changes that were happening in the prokaryotic community alone.

In NMDS analysis, a clear split between host-depleted samples and the control appeared: This split appeared along the first axis. All host depleted samples were clustering on one end, while the control samples clustered on the other. The distance between the two sets was not very large, the spread inside each set was far bigger. There was also clear overlap between the 95% confidence intervals of the samples depleted with the HostZero kit and those extracted with the PowerSoil kit. No samples fell into this area, but there could have been theoretical samples fitting both. There was no overlap between the samples depleted with the QIAamp kit and the samples extracted with the PowerSoil kit (Figure 8.).

I considered that this difference could be due to the difference in host organelle content. After all, having significantly fewer of the most abundant groups of reads might have been a large difference. This would have been a good sign, because their removal is the goal of host depletion. To see if this was the case, I removed all mitochondrial and chloroplast reads from the dataset *in-silico*, both from the depleted samples and from the control. Afterwards, I repeated NMDS with the exact same parameters as before.

The difference between the two sets of samples did not vanish when removing the host organelles, but it decreased. There was now overlap between all treatments, with many samples of all treatments in this area of overlap. About a third of all samples depleted with either the QIAamp or the HostZero kit fell into the confidence interval of the PowerSoil kit, and even a few isolated with the PowerSoil kit fells into the confidence interval of both the HostZero and the QIAamp kits. The shape of the confidence intervals of the two host depletion kits also changed, it became more stretched along the first axis, which lost the property of fully separating them from the control. This was no longer possible along any of the axes. The distance between the two sets of samples also decreased (Figure 8.).

The difference between the samples was still highly significant. The difference between the samples extracted with the PowerSoil kit on the one hand, and the samples extracted both with the HostZero and the QIAamp kit on the other still had almost the same significance as before ($p=3x10^{-4}$ for both QIAamp and HostZero). The explanatory power of that difference, the R^2, did decrease, to about 0.15, less than half its previous value (Figure 8.).

These results show that host depletion introduces biases into the sequenced seagrass leaf microbiome. This further confirmed the data from the Venn diagram analysis. Host depletion does not only lead to many taxa missing from the sequencing data, but it also introduces biases into the whole structure of the prokaryotic community that is observed through sequencing.

Host depletion kits deplete archaeal abundance along with host 'contamination'

Archaeal cells are more similar to eukaryotic cells than to bacterial cells. I therefore wanted to investigate the effect of host depletion kits on the abundance of archaeal reads in the samples. It seemed possible that host depletion kits also reduce the abundance of archaeal reads. This would be another source of bias introduced into the sequencing results. I evaluated the changes in archaeal reads between the different treatments.

Archaeal reads made up quite a small percentage of the sequenced reads. On average, just 2.3% of all reads in the samples that were extracted with the PowerSoil were identified as archaeal. But there were large differences between this and the host depletion kits.

All host depletion kits also depleted archaea. In samples depleted with the QIAamp kit, archaea made up only 0.8% of the sequencing reads, a statistically significant (p=0.0167) decrease by 65.2% compared to the control. The HostZero kit reduced archaeal reads to just 0.6% of the sequencing reads, a statistically significant(p=0.0091) decrease of 73.9% compared to the conventional DNA extraction (Figure 6.).

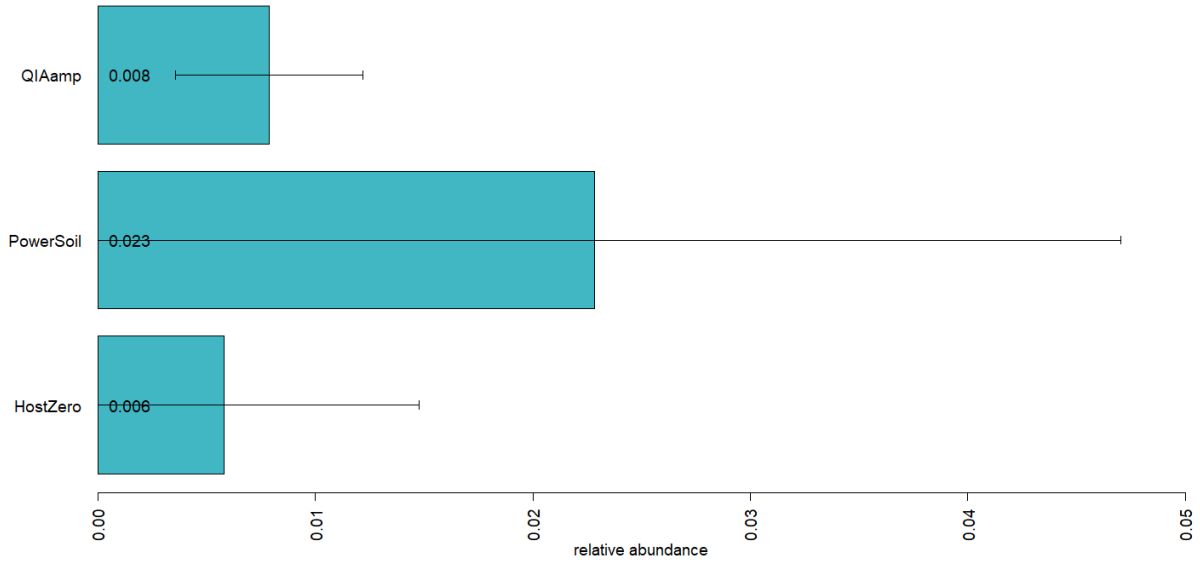


Figure 10. ***Average archaeal content of samples treated with different host depletion kits.*** *The x-axis depicts the relative abundance in the different samples sequenced with and without PNAs. The y-axis lists the method that was used for host depletion. The numbers in each bar signify the average value for that sample. The relative archaeal content did significantly decrease with both HostZero (p=0.0091) and QIAamp (p=0.0167).*

Host depletion kits not only introduced biases into the microbial community, they also selectively reduced the relative abundance of archaeal reads in the sample. This again added to the fact that the sequencing results gained from samples where host depletion has been performed were biased. Therefore, every analysis of the prokaryotic community of seagrass leaves done on samples depleted with these kits would likely be missing some parts.

Host depletion enabled deeper analysis of the seagrass microbiome

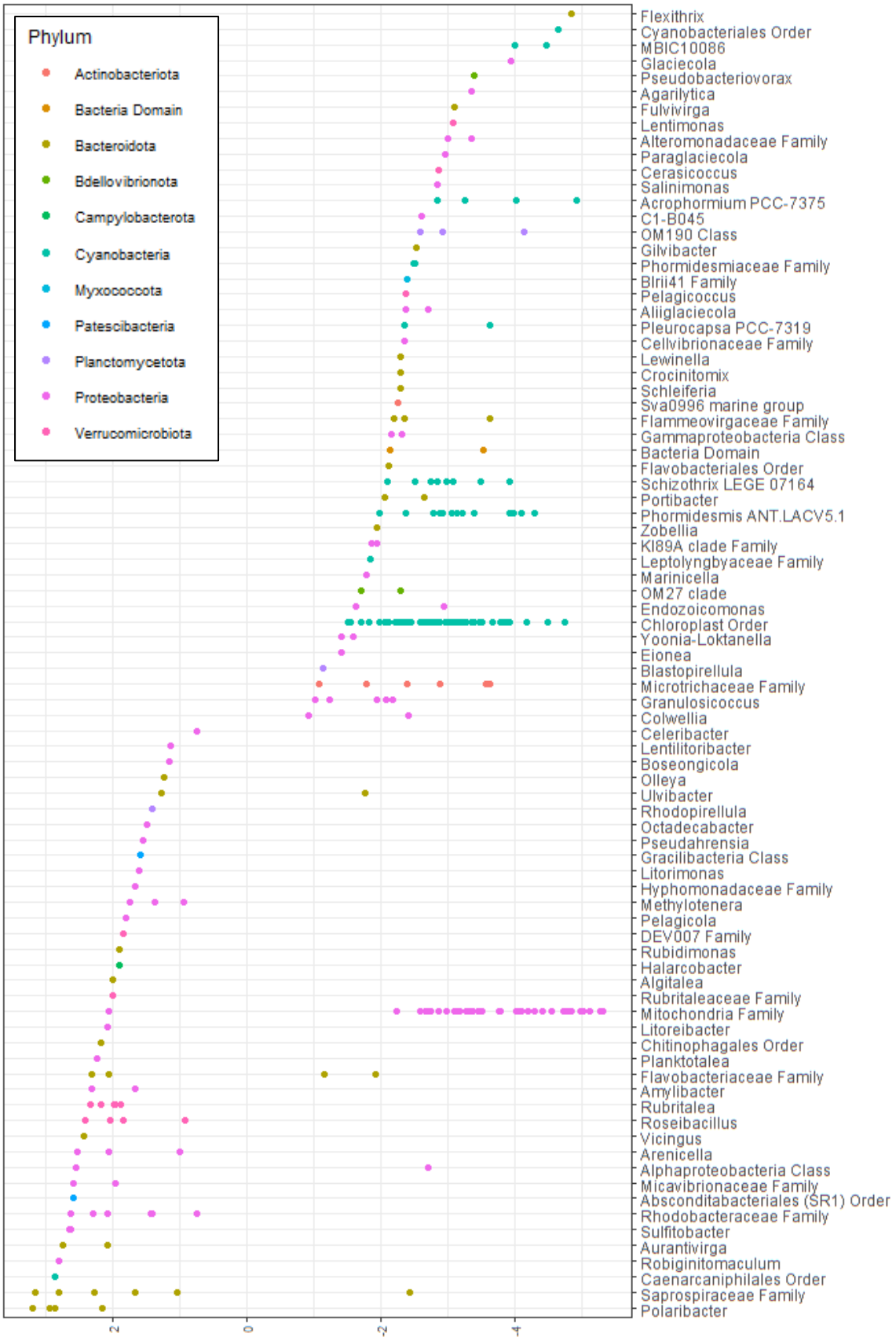


Figure 11. ***Changes in the abundance of genera with host depletion.*** *The figure depicts the fold-change of genera between samples depleted with the QIAamp kit and samples extracted with the PowerSoil kit. The x-axis depicts the base-2 logarithm of the fold-change of the genera listed on the y-axis. The y-axis lists all genera where the fold-change is statistically significant. Each dot represents a single species of that genus. The colours of the dots denote the assignment of the genus to its phylum.*

I also wanted to investigate how specific genera are behaving as a result of host depletion. I wanted to see whether they are increasing or decreasing in abundance. For this, I needed to calculate the fold-change of each genus. Here, I decided to use the fold-change between the samples depleted with the QIAamp kit, the more successful of the two host depletion kits, and the samples extracted with the PowerSoil kit, the control.

It was clearly visible that host depletion with the QIAamp kit reduced the host organelle content. The highest fold change for mitochondria was still below 0.25, a decrease by 75% due to host depletion. The lowest was at 0.022, a decrease by more than 98%. The average fold change was 0,082, a decrease by more than 92%. For chloroplasts, this ranged from 0.35, a decrease by 65% due to host depletion, to 0.031, a decrease by more than 97%. The average fold change here was 0.129, a decrease by 87% (Figure 10.).

It became obvious that some other ASVs at the genus level were also strongly and significantly depleted by the QIAamp kit. This included ASVs assigned to the genera *Flexithrix and Glaciecola*, both of which were decreasing more strongly on average than even mitochondria. There were also an additional eight ASVs at the genus level that were decreased less strongly on average than mitochondria, but still more strongly than chloroplasts (they were assigned to the genera *Pseudobacteriovorax*, *Agarilytica*, *Phormidesmis*, *Fulvivirga*, *Lentimonas*, *Pleurocapsa*, *Schizothrix*, and *Paraglaciecola*; Figure 10.).

Other ASVs at the genus level were significantly increased in abundance. ASVs of different genera were increased strongly in the samples depleted with the QIAamp kit. This increase was very strong for some of them, up to almost 9-fold. This was true for different ASVs with no genus assigned in the family Saprospiraceae, some ASVs with no assigned family or species in the order Caenarcaniphilales, and for ASVs assigned to the genera *Sulfitobacter*, *Polaribacter*, *Aurantivirga*, *Amylibacter*, and *Robiginitomaculum* (Figure 10.).

It was visible that more ASVs were significantly decreased than increased in abundance. This was true, even when ASVs assigned to mitochondria and chloroplasts were disregarded. The range of fold-change in the significantly decreased ASVs was also larger, going from -1 to -5, while it only spanned +1 to +3 for the significantly increased ASVs (Figure 10.).

These results again hint at the biased nature of host depletion. Many genera are being reduced in relative abundance by host depletion, some of them even stronger than the host organelles. At the same time, other taxa are massively increased in relative abundance. These taxa might have been missed without host depletion.

Discussion

The QIAamp kit worked better than HostZero for host depletion in seagrass, but both kits still have their limitations

In this study, I wanted to identify the best method to deplete host 'contamination' in seagrass. I compared different approaches to do that, both kit-based and non-kit-based. The kit-based approach was the only one that significantly and strongly reduced host contamination. The kit that worked best was the QIAamp kit from Qiagen, a pre-extraction host depletion kit.

Pre-extraction kits like QIAamp are the most widespread form of host depletion kits. Most commercially available kits follow this procedure; and many, non-kit-based, approaches in literature are variations on this approach [48, 50, 61, 63, 86, 87]. But these kits are also more labour-intensive and require larger amounts of biomass than any other approach [60].

In my study, the QIAamp kit was the best kit for host depletion in seagrass leaves. QIAamp reduced the host content by about half. In this, it was closely followed by the HostZero kit, which reduced it by about one third (Figure 6.). The QIAamp kit did also significantly increase the number of observed prokaryotic taxa in the sequenced samples (Figure 7. and Supplementary Figure 4.). The HostZero kit did not achieve that. But there are also results showing that host depletion introduces biases into the sequencing results.

In the samples sequenced with each of these kits, there were different genera detected than in the other kit and the control. Each kit missed as much as a fifth of all the taxa detected. Almost a tenth of all taxa were not detected by any of the host depletion kits (Figure 8.). This shows that both host depletion kits were not able to capture the full spectrum of organisms detected in the community.

The community composition was changed significantly by both kits compared to the control. These changes were not explained by the removal of host organelles from the sample. The *in-silico* removal of the host organelles from the dataset reduced the differences between the different treatments and the control, but the difference was still significant. Therefore, sequencing the DNA extracted and depleted with these kits from the seagrass leaves will not present the same community as detected after 'normal' DNA extraction.

These results taken together gave a divided picture on the usefulness of host depletion kits to study the DNA. On the one hand, the kits were undeniably successful in depleting host organelles. They significantly reduced host 'contamination' measured by the relative abundance of host organelles in the sequenced dataset. The QIAamp kit also increased the observed diversity. On the other hand, it also cannot be denied that both kits introduced biases into the community. They both missed parts of the community completely. In addition, many bacterial genera decreased in relative abundance due to host depletion. So, all together, these kits might not be ideal to study the prokaryotic community of seagrass leaves.

But both kits are still useful for analysis of the seagrass leaf microbiome. While quantitative community studies might not be possible with these kits, other studies are only made possible by their use. Host depletion with the QIAamp kit showed 400 more bacterial species in our sample than 'normal' DNA extraction. For any study aiming to find specific organisms or to identify as many members of the community as possible, this approach would work perfectly.

Host depletion kits deplete archaeal sequences from the seagrass leaf microbiome

Another bias introduced into the data by host depletion is the depletion of archaeal sequences. Both kits significantly deplete archaeal reads in the sequencing results, QIAamp by two thirds, HostZero even more, by three quarters (Figure 10.). This is likely due to the method used to differentiate between prokaryotic and eukaryotic cells by these kits. They are selectively lysing eukaryotic cells, and it seems that archaeal cells are more like eukaryotic than prokaryotic ones. It is interesting, though, that the kit that is more successful at depleting eukaryotic cells is slightly less successful at depleting archaeal cells.

Archaea were present in very low abundance in my samples. All archaea taken together made up little more than 2% of the reads sequenced. Archaeal abundance was so low that no real statement about 'archaeal depletion' could be made with any certainty from my data alone.

There are currently no systematic studies available on the effect of host depletion on archaeal abundance. One study that is currently only available as a preprint looked at the effect of different host depletion methods on an artificial community containing one archaeal species. They had no conclusions regarding the archaeon, as they did not have enough sequencing depth to detect it even in the non-host-depleted control [88]. Another study did not compare host depletion methods, but used a specific pre-extraction host depletion method they had designed themselves. They were able to assemble a full genome of the archaeon *Methanobrevibacter smithii* from their host depleted samples [89]. This is coincidentally the very same archaeon that could not be detected in the preprint mentioned before [88]. Therefore, it seems possible to use host depletion in a way that does not deplete archaea.

It is unclear how exactly host depletion can be performed without depleting archaea. In their study, the authors used an individually designed host depletion protocol. This protocol seems very similar to the protocols of both host depletion kits used in this study. The only clear differences in their protocol is that they homogenized their samples using a FastPrep for bead beating, and avoided high-temperature inactivation steps [89]. It would be necessary to test which of these changes has the intended effect of maintaining relative archaeal abundance in the samples.

Host depletion kits deplete only mitochondrial, not chloroplast sequences

I noticed one curious aspect of host depletion: Both host depletion kits successfully reduced relative host organelle content and relative mitochondrial abundance. At the same time, neither of the kits was able to deplete chloroplast sequences from the samples at all. Relative chloroplast abundance stayed the same upon treatment with the QIAamp kit, and even increased upon treatment with the HostZero kit (Figure 6.). This difference in the kits' effectiveness was very surprising.

This discrepancy might be due to what the host depletion kits are designed for. Both host depletion kits used in this study are designed for human samples. For both of them, the product pages declares that they are intended for 'eukaryotic samples', with no further limitations. But as the only examples for samples that can be used, the manufacturer gives bodily fluids, swabs, and saliva [90, 91]. These are all kinds of samples that would be gathered in a medical setting, from humans. This clearly suggests that plant samples are not what this kit was designed for.

But even if this kit is not designed for plant samples, it can still be useful. Plants have mitochondria, too, and they make up a larger portion of the sequenced reads than chloroplasts. Every kit that can

significantly deplete this part of the host organelles is already reducing host 'contamination'. Even if chloroplasts cannot be removed at all, this still frees up sequencing power for the microbiome.

The used peptide nucleic acid clamps do not effectively block sequencing of *Z. marina* organelles

PNAs were only added in the last step, right before sequencing the samples, and only to samples extracted with the PowerSoil kit. It was clear that sequencing with PNAs did not have the intended effect on the host content observed in the sequencing results. There were small differences in the chloroplast content, but no clear reduction of mitochondria (Figure 4.).

This was very strange, as 16S rRNA gene sequencing would be the obvious application of PNAs. This is because their ability to effectively deplete specific sequences in library preparation would make them an ideal choice. The manufacturer explicitly recommends this use [92]. This is also something PNAs have been used for in other studies [93–95].

But the manufacturer also states that it might be possible that their standard PNAs do not bind to all mitochondria and chloroplasts, in which case custom-made PNAs would be required [92]. And indeed, when I BLASTed the sequences of the mPNA against the reference genome for the mitochondrion of *Zostera marina* [96], and the sequence of the pPNA against the reference genome for the chloroplast of *Zostera marina* [97], I found that there is no (perfect) match. In future, this approach may be more successful with custom-designed oligos for the organelle sequences from these target species.

This is the first systematic study of host depletion in plants

This study is the first of its kind. There are very few studies using host depletion kits available at all. There are less studies comparing them to 'normal' DNA extraction. There are even less studies comparing different host depletion kits. There is only one single study on host depletion in plant microbiomes, and even this was done on root samples, not leaves. There are no studies on host depletion in environmental samples.

The QIAamp kit has not been used in many studies. Currently (2024/09/21), there are slightly more than 500 available on Google Scholar, almost 100 of these published this year alone. But the vast majority of these studies are on human or mammal samples. There are some studies on the microbiome of many different organisms where the QIAamp kit is used as the only kit. Species studied in those studies include *Artemia* [98], the fish *Oreochromis niloticus* [99], *Centropomus undecimalis* [100], *Alburnus derjugini*, *Salmo labrax*, and *Oncorhynchus mykiss* [101], the snake *Protobothrops mucrosquamatus* [102], *Daphnia* [103], the microalgae *Alexandrium tamarense* and *Cochlodinium polykrikoides* [104], as well as different insect species [105–110]. There is also one single study on plant microbiomes using the QIAamp kit. This study is on the microbiome of the roots of the arecanut tree [111]. In this study, the authors are only using this kit, so no comparison to validate host depletion is available, but their results suggest that the kit works also for studying a plant microbiome.

The HostZero kit has been used in even fewer studies. There are currently (2024/09/21) only about 50 studies available on Google Scholar that use it. Most of these, again, are on human or mammal samples. There are just a few studies investigating its usefulness in the study of animal microbiomes, including in the coconut beetle *Brontispa longissimi* [112], the nematode *Caenorhabditis elegans* [113], the coral *Galaxea fascicularis* [114], and salmon [115]. There is also one study on chlamydia infecting

the photosynthetic dinoflagellate *Cladocopium* sp., a symbiont of corals [116]. There is only one study where the HostZero kit is used on plant material. This study is from the field of food safety, so the only 'plant material', mashed potatoes, is highly processed [58]. In all these studies, HostZero is used as the only kit, so there is no comparison to a non-host depleting kit. This means the efficacy of host depletion by the HostZero kit in any of these samples cannot be estimated.

This is not the case for one other study. It is a host depletion study on the microbiome of a bivalve, *Perna canaliculus* [117]. In this study, they compared the HostZero kit with the PowerSoil kit for DNA extraction, showing HostZero to be extremely effective in depleting host tissue [117]. This study is also of interest for the study of the ecology of seagrass, as this species sometimes occurs in seagrass meadows [118].

There are currently very few studies comparing the effect of host depletion kits with each other on the same sample type. There are currently (2024/09/20) less than 10 different studies available on Google Scholar (Table 1.). Some of these evaluated other kits than used in this study. All of them are on human tissue samples. These studies reach very different results.

The results of these studies varied a lot. In biopsies from diabetic foot infections, HostZero was slightly better than QIAamp. In urine samples, HostZero was far better than QIAamp, while QIAamp barely worked. In intestinal biopsies, none of the kits worked extremely well, but QIAamp clearly worked better than HostZero (Table 1.). It is almost impossible to draw any general conclusions which kit is better. It is also impossible to predict which kit will perform best on any sample type on which it has not yet been tested.

Table 1. ***Comparison of host depletion kits in this and other studies.*** *The table lists all (to my knowledge) available, comparative studies of host depletion that use the two kits used in this study. It shows the relative depletion of host reads by each kit compared to the control used in the respective study and lists the sample type the study was performed on, as well the reference for that study.*

Sample type	**Depletion with QIAamp**	**Depletion with HostZero**	**Reference**
Biopsy samples from severe diabetic foot infections	-69.0%	-78.5%	[50]
Human intestinal biopsies	-27.2%	-6.1%	[61]
Urine samples	-4.9%	-63.4%	[88]
Seagrass leaves	-65.7%	-32.8%	This study

This makes this study the first of its kind in many ways. It is the first study of host depletion in leaf tissue, and the first host depletion study on marine plants. This study is also the first comparative study on host depletion in plants, and also in environmental samples more general. It is even the first comparative study of host depletion done on non-human samples. It showed that host depletion can be used outside of human-centred, medical studies. It can also be used for plant samples, and not only in tissue samples, but in environmental samples, as well.

The microbial community detected in this study was consistent with previous studies of the *Zostera marina* leaf microbiome

I analysed and identified the most abundant families of the microbiome. I wanted to compare this with other published studies giving the species composition of the *Z. marina* leaf microbiome. The prokaryotic community was dominated by four families. These families were Rhodobacteraceae, Flavobacteriaceae, Phormidesmiaceae, and Rubritaleaceae.

Other studies often only give taxonomical assignments to higher levels than the genus level. Nevertheless, comparisons were still possible. The dominant taxa in other studies often were the Proteobacteria. These taxa made up the largest group, between 35 and 70% of all ASVs were assigned to them. The second largest phylum were the Bacteroidota, making up between 10 and 20% of the community [41, 119, 120]. A family that was observed in very large numbers

This matched the results from our community very well. Of all reads sequenced, Proteobacteria were the largest phylum, including the largest family, the Rhodobacteraceae. The second phylum, the Bacteroidota, was represented by the Flavobacteriaceae, the second-largest family, as well as by the Saprospiraceae, the fifth-largest family in my samples.

But the dominant representatives in my sequencing results were, of course, the host organelles. Even if the community structure is still visible in 'normal' 16S rRNA gene amplicon sequencing, it is still drowned out by host 'contamination'. This might obscure details of the community. In addition, it would make smaller changes invisible in comparative studies. Therefore, host depletion is a useful tool for the analysis of the seagrass microbiome.

Host depletion was able to detect the core seagrass microbiome

The microbiome of *Zostera marina* leaves was described to contain just a few taxa. These bacteria are consistently present in *Z. marina* plants, irrespective of sampling site. Some of them are present on the plant even when they are not found in the surrounding seawater. Of these 11 OTUs, six belong to the genus *Methylotenera*. The others are one OTU each of the genera *Paraglaciecola*, *Alteromonas*, *Marinimonas*, *Rubidimonas*, and one OTU is an unidentified member of the family Rhodobacteraceae [40, 41]. All of these genera were detected in my samples after host depletion.

Host depletion had different effects on the detection of these core genera. Both *Methylotenera* and *Rubidimonas* increased significantly in relative abundance. *Paraglaciecola* were depleted strongly, while *Alteromonas* and *Marinimonas* did not change in abundance due to host depletion. This means that host depletion would enable deeper sequencing and deeper study of seven of the eleven core microbiota of *Z. marina* leaves.

Seagrass leaves exude methanol [121]. In this, seagrasses are similar to land plants. Methanol is toxic to plants, as it can inhibit germination and seedling growth [122]. In some land plants, this problem is solved by the presence of a bacterium, a member of the genus *Methylobacterium*. This bacterium consumes the produced methanol. While *Methylobacterium* is not found in the seagrass, many other methanol-consuming bacteria are found [122, 123].

Methylotenera use single-carbon sources, including methanol, for their metabolism [121]. Because of this, it has been suggested that they might play a similar role to *Methylobacterium* in land plants [40, 123]. To be able to better study the potential relationship between seagrass leaves and *Methylotenera*,

more studies would be necessary. Using host depletion would make this study easier, especially as it has been shown to increase the detected relative abundance of the organism.

Host depletion enabled deeper analysis of the seagrass microbiome

Many bacterial genera increased strongly after host depletion (Figure 11.). I investigated the three genera that increased the strongest. Species of those genera might very well have been missed if not for host depletion. Those genera were *Polaribacter*, *Robiginitomaculum*, and *Sulfitobacter*. While there were species not belonging to these taxa that increased more strongly, it was not possible to taxonomically assign them to a genus, so I decided to disregard them for this analysis.

Polaribacter is a genus in the Flavobacteriaceae with 29 species described. They are heterotrophic aerobes [124]. They are often found associated to different kinds of macroalgae [125–129]. They are specialised in degrading alginates and fucoidans, major components of brown algae [130, 131]. They are known as degraders of kelp tissue [132], which is likely the reason why they are increasingly found on kelp that is suffering due to increased water temperature [133]. It has been shown that different clades of *Polaribacter* are capable of utilising different algal compounds, leading to niche differentiation [134]. Species of this genus have already been detected on seagrass leaves [135]. It has even been suggested that they might provide the nutrients they derive from brown algae degradation to the seagrass they inhabit, although no proof of that has been observed so far [132].

Robiginitomaculum is a genus in the Hyphomonadaceae. It contains just one species, *Robiginitomaculum antarcticum*, which is a heterotrophic aerobe capable of nitrate reduction. It was isolated originally from the water column off the Antarctic coast [136]. Like *Polaribacter*, it is often found on kelp and other algae [137–139]. Little else is known about it. To the best of my knowledge, there is no mention of any connection to seagrass. But it stands to reason that it could serve a similar ecological role to *Polaribacter* in the seagrass microbiome.

Sulfitobacter is a genus in the Roseobacteraceae with 25 described species. They are heterotrophic aerobes and can oxidize both sulphur and sulphite to sulphate [140]. They are very abundant in seawater [141], but are also found associated with different micro-algae [142, 143]. They interact with the algae that harbour them in various ways, often contributing to the regulation of algal blooms. The species *Sulfitobacter pontiacus* protects the micro-algae *Emiliania huxleyi* against the algicidal bacterium *Phaeobacter inhibens*, which can cause the death of the algae [143]. A different Sulfitobacter species, associated with the same species of micro-algae, can utilise dimethylsulfoniopropionate, DMSP, a product of the algal metabolism. It produces methanethiol, which is toxic to the algae and increases their susceptibility to algicidal bacteria [144]. Another species, *Sulfitobacter porphyrae*, also exhibits an algicidal capability. In the presence of the micro-algae *Prorocentrum donghaiense*, it is capable of producing algicidal compounds that kill the algae, as well as other algal species, *Alexandrium minutum* and *Scrippsiella trochoidea*. All three of these species are dangerous because of their harmful, toxic blooms [145]. Nothing is known about potential interactions of *Sulfitobacter* species and seagrass, but two possible interactions of seagrass, *Sulfitobacter*, and algae appear possible. They also might be protecting seagrass against bacterial pathogens. But another thing seems even more likely: Given their strong interactions with micro-algae and algal blooms, it could be possible that they provide some kind of protection against harmful algal blooms to seagrass.

Sulfitobacter are known as sulphur reducers [140]. This suggests another potential role of theirs in the seagrass leaf microbiome: In the root microbiome of seagrass, a different family of sulphate reducers,

Sedimenticolaceae, is present. It has been described as extremely important for the existence of seagrass ecosystems by detoxifying the surrounding sediment [29]. It seems possible that *Sulfitobacter* might play a similar role on seagrass leaves.

All in all, not much is known about possible interactions between these three genera of bacteria and seagrass. Nevertheless, they play important roles in marine ecology, and their possible association or interaction with seagrass might be important. Every such interaction that might be postulated would of course need to be investigated and evaluated. But to be able to postulate and then investigate any interaction, it is necessary to know about the presence of the partners of the interaction. This is exactly how host depletion could be an extremely useful tool. Each of these species might easily have been missed in a study where the seagrass microbiome was sequenced far less deeply.

Conclusions

With this study, I made the first comparative study of host depletion kits in seagrass leaves for 16S rRNA gene amplicon sequencing. This is also the first study of its kind on environmental and plant samples.

The microbiome, as sequenced without host depletion, was dominated by host organelles. They made up the large majority of all reads sequenced. Even though the prokaryotic community structure was still visible, host depletion would make community analysis far easier.

I compared different methods to deplete host content in seagrass leaves for 16S rRNA gene amplicon sequencing. The best method to achieve that for these samples was the QIAamp Microbiome Kit from Qiagen, with an additional filtration step added at the beginning. While this kit did not extract the community composition as it is in the environment, its large effect on the diversity and the potential to discover rarer species in the environment still make it an important tool for studying the seagrass leaf microbiome.

With host depletion, it was possible to detect previously described core microbiota of seagrass in our samples. These taxa might play an important role in seagrass ecology. Many genera from the seagrass leaf microbiome are increasing strongly with host depletion. Without host depletion, they might have been missed. Some of them are known to have ecological and metabolic functions that could be important in interactions with seagrass. To study these low-abundance organisms, host depletion would be an invaluable tool.

Outlook

The possible uses of host depletion methods in environmental microbiology or microbial ecology are virtually endless. Every host microbiome, every symbiotic community, every ecosystem containing eukaryotes, all that could be sequenced far deeper. Even as sequencing becomes more readily available, this approach can still free up two thirds or more of each sequencing run for sequencing microbes, not eukaryotes that are already well-known.

To study the seagrass leaf microbiome as a community using host depletion, more research is needed to identify a way to use host depletion to gain an unbiased picture of the community. But even without that, host depletion can still be useful. It would be possible to study the core microbiota in more depth. In addition, it would also be interesting to investigate further the non-core microbiome. This might make it possible to identify the factors leading to the variation of the seagrass microbiome.

The obvious next step would be to apply host depletion to metagenomics. In this study, I showed the effect host depletion has on seagrass in 16S rRNA gene amplicon sequencing. But both kits I used can also be utilised to deplete host ‘contamination’ for metagenomics. Studying their effect there would be an important next step.

In general, it should be possible to use host depletion for more comprehensive studies of changes in the microbiome of seagrass. These studies could investigate many different things: the changes in the microbiome of one plant during the course of the seasons, differences between plants of the same species at different sites, differences between different species of seagrass at the same site, or the effects of various environmental stressors

References

1. Ettinger CL, Eisen JA. Fungi, bacteria and oomycota opportunistically isolated from the seagrass, *Zostera marina*. *PLoS ONE* 2020; **15**: e0236135.
2. Short F, Carruthers T, Dennison W, Waycott M. Global seagrass distribution and diversity: a bioregional model. *Journal of Experimental Marine Biology and Ecology* 2007; **350**: 3–20.
3. The Angiosperm Phylogeny Group, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 2016; **181**: 1–20.
4. Duarte B, Repolho T, Paula JR, Caçador I, Matos AR, Rosa R. Ocean acidification alleviates dwarf eelgrass (*Zostera noltii*) lipid landscape remodeling under warming stress. *Biology* 2022; **11**: 780.
5. Prazukin AV, Firsov YK, Latushkin AA. The *Zostera Noltii* Hornemann seagrass as the modulator of habitat within its canopy. 2024. SSRN, Rochester, NY.
6. Benmokhtar S, Robin M, Maanan M, Boutoumit S, Badaoui B, Bazairi H. Monitoring the spatial and interannual dynamic of *Zostera noltei*. *Wetlands* 2023; **43**: 43.
7. Jones KL, Hartl MGJ, Bell MC, Capper A. Microplastic accumulation in a *Zostera marina* L. bed at Deerness Sound, Orkney, Scotland. *Marine Pollution Bulletin* 2020; **152**: 110883.
8. McKenzie LJ, Nordlund LM, Jones BL, Cullen-Unsworth LC, Roelfsema C, Unsworth RKF. The global distribution of seagrass meadows. *Environmental Research Letters* 2020; **15**: 074041.
9. Nordlund LM, Unsworth RKF, Gullström M, Cullen-Unsworth LC. Global significance of seagrass fishery activity. *Fish and Fisheries* 2018; **19**: 399–412.
10. Duarte CM. The future of seagrass meadows. *Environmental Conservation* 2002; **29**: 192–206.
11. Maxwell PS, Eklöf JS, van Katwijk M, O'Brien K, de la Torre-Castro M, Boström C, et al. The fundamental role of ecological feedback mechanisms for the adaptive management of seagrass ecosystems – a review. *Biological Reviews* 2017; **92**: 1521-1538.
12. Panno L, Bruno M, Voyron S, Anastasi A, Gnavi G, Miserere L, et al. Diversity, ecological role and potential biotechnological applications of marine fungi associated to the seagrass *Posidonia oceanica*. *New Biotechnology* 2013; **30**: 685–694.
13. Mohr W, Lehnen N, Ahmerkamp S, Marchant HK, Graf JS, Tschitschko B, et al. Terrestrial-type nitrogen-fixing symbiosis between seagrass and a marine bacterium. *Nature* 2021; **600**: 105–109.
14. Cullen-Unsworth LC, Nordlund LM, Paddock J, Baker S, McKenzie LJ, Unsworth RKF. Seagrass meadows globally as a coupled social–ecological system: implications for human wellbeing. *Marine Pollution Bulletin* 2014; **83**: 387–397.
15. Ugarelli K, Chakrabarti S, Laas P, Stingl U. The seagrass holobiont and its microbiome. *Microorganisms* 2017; **5**: 81.
16. Lamb JB, van de Water JAJM, Bourne DG, Altier C, Hein MY, Fiorenza EA, et al. Seagrass ecosystems reduce exposure to bacterial pathogens of humans, fishes, and invertebrates. *Science* 2017; **6326**: 731-733.
17. Kannan RRR, Arumugam R, Anantharaman P. Antibacterial potential of three seagrasses against human pathogens. *Asian Pacific Journal of Tropical Medicine* 2010; **3**: 890–893.

18. Watson RA, Coles RG, Long WL. Simulation estimates of annual yield and landed value for commercial penaeid prawns from a tropical seagrass habitat, Northern Queensland, Australia. *Marine and Freshwater Research* 1993; **44**: 211–219.

19. Hughes AR, Williams SL, Duarte CM, Heck Jr KL, Waycott M. Associations of concern: declining seagrasses and threatened dependent species. *Frontiers in Ecology and the Environment* 2009; **7**: 242–246.

20. Orth RJ, Carruthers TJB, Dennison WC, Duarte CM, Fourqurean JW, Heck KL, et al. A global crisis for seagrass ecosystems. *BioScience* 2006; **56**: 987–996.

21. Moriarty DJW, Iverson RL, Pollard PC. Exudation of organic carbon by the seagrass *Halodule wrightii* Aschers and its effect on bacterial growth in the sediment. *Journal of Experimental Marine Biology and Ecology* 1986; **96**: 115–126.

22. Holmer M, Duarte CM, Boschker HTS, Barrón C. Carbon cycling and bacterial carbon sources in pristine and impacted Mediterranean seagrass sediments. *Aquatic Microbial Ecology* 2004; **36**: 227–237.

23. Kirchman DL, Mazzella L, Alberte RS, Mitchell R. Epiphytic bacterial production on *Zostera marina*. *Marine Ecology Progress Series* 1984; **15**: 117–123.

24. Patriquin D, Knowles R. Nitrogen fixation in the rhizosphere of marine angiosperms. *Marine Biology* 1972; **16**: 49–58.

25. Agawin NSR, Ferriol P, Cryer C, Alcon E, Busquets A, Sintes E, et al. Significant nitrogen fixation activity associated with the phyllosphere of Mediterranean seagrass *Posidonia oceanica*: first report. *Marine Ecology Progress Series* 2016; **551**: 53–62.

26. Hamisi MI, Lyimo TJ, Muruke MHS, Bergman B. Nitrogen fixation by epiphytic and epibenthic diazotrophs associated with seagrass meadows along the Tanzanian coast, Western Indian Ocean. *Aquatic Microbial Ecology* 2009; **57**: 33–42.

27. van der Heide T, Govers LL, de Fouw J, Olff H, van der Geest M, van Katwijk MM, et al. A three-stage symbiosis forms the foundation of seagrass ecosystems. *Science* 2012; **336**: 1432–1434.

28. Cardini U, Marín-Guirao L, Montilla LM, Marzocchi U, Chiavarini S, Rimauro J, et al. Nested interactions between chemosynthetic lucinid bivalves and seagrass promote ecosystem functioning in contaminated sediments. *Frontiers in Plant Science* 2022; **13**.

29. Martin BC, Middleton JA, Fraser MW, Marshall IPG, Scholz VV, Hausl B, et al. Cutting out the middle clam: lucinid endosymbiotic bacteria are also associated with seagrass roots worldwide. *The ISME Journal* 2020; **14**: 2901–2905.

30. Tarquinio F, Hyndes GA, Laverock B, Koenders A, Säwström C. The seagrass holobiont: understanding seagrass-bacteria interactions and their role in seagrass ecosystem functioning. *FEMS Microbiology Letters* 2019; **366**: fnz057.

31. Armstrong E, Yan L, Boyd KG, Wright PC, Burgess JG. The symbiotic role of marine microbes on living surfaces. *Hydrobiologia* 2001; **461**: 37–40.

32. Lucas-Elio P, Hernandez P, Sanchez-Amat A, Solano F. Purification and partial characterization of marinocine, a new broad-spectrum antibacterial protein produced by *Marinomonas mediterranea*. *Biochimica et Biophysica Acta* 2005; **1721**: 193–203.

33. Lucas-Elío P, Gómez D, Solano F, Sanchez-Amat A. The antimicrobial activity of marinocine, aynthesized by *Marinomonas mediterranea*, is due to hydrogen peroxide generated by its lysine oxidase activity. *Journal of Bacteriology* 2006; **188**: 2493–2501.

34. Ravikumar S, Gnanadesigan M, Saravanan A, Monisha N, Brindha V, Muthumari S. Antagonistic properties of seagrass associated *Streptomyces* sp. RAUACT-1: a source for anthraquinone rich compound. *Asian Pacific Journal of Tropical Medicine* 2012; **5**: 887–890.

35. Burja AM, Banaigs B, Abou-Mansour E, Grant Burgess J, Wright PC. Marine cyanobacteria—a prolific source of natural products. *Tetrahedron* 2001; **57**: 9347–9377.

36. Mazard S, Penesyan A, Ostrowski M, Paulsen IT, Egan S. Tiny microbes with a big impact: the role of cyanobacteria and their metabolites in shaping our future. *Marine Drugs* 2016; **14**: 97.

37. Celdrán D, Espinosa E, Sánchez-Amat A, Marín A. Effects of epibiotic bacteria on leaf growth and epiphytes of the seagrass *Posidonia oceanica*. *Marine Ecology Progress Series* 2012; **456**: 21–27.

38. Jankowska E, Jankowska K, Włodarska-Kowalczuk M. Seagrass vegetation and meiofauna enhance the bacterial abundance in the Baltic Sea sediments (Puck Bay). *Environmental Science and Pollution Research International* 2015; **22**: 14372–14378.

39. Fahimipour AK, Kardish MR, Lang JM, Green JL, Eisen JA, Stachowicz JJ. Global-scale structure of the eelgrass microbiome. *Applied and Environmental Microbiology* 2017; **83**: e03391-16.

40. Sanders-Smith R, Segovia BT, Forbes C, Hessing-Lewis M, Morien E, Lemay MA, et al. Host-specificity and core taxa of seagrass leaf microbiome identified across tissue age and geographical regions. *Frontiers in Ecology and Evolution* 2020; **8**.

41. Adamczyk EM, O'Connor MI, Parfrey LW. Seagrass (*Zostera marina*) transplant experiment reveals core microbiota and resistance to environmental change. *Molecular Ecology* 2022; **31**: 5107–5123.

42. Ma X, Vanneste S, Chang J, Ambrosino L, Barry K, Bayer T, et al. Seagrass genomes reveal ancient polyploidy and adaptations to the marine environment. *Nature Plants* 2024; **10**: 240–255.

43. Cui K, Wang S, Pei Y, Zhou B. Occurrence and distribution of antibiotic pollution and antibiotic resistance genes in seagrass meadow sediments based on metagenomics. *Science of The Total Environment* 2024; **935**: 173438.

44. Mohapatra M, Dash SP, Rastogi G. Genomic approaches for the investigation of seagrass rhizosphere microbiome and bioprospecting potential: a field study from Chilika Lagoon, Odisha, India. In: Thatoi H, Pradhan SK, Kumar U (eds). *Applications of Metagenomics*. 2024. Academic Press, pp 201–234.

45. Abdel-Wahab MA, Bahkali AH, Elgorban AM, Jones EBG. High-throughput amplicon sequencing of fungi and microbial eukaryotes associated with the seagrass *Halophila stipulacea* (Forssk.) Asch. from Al-Leith mangroves, Saudi Arabia. *Mycological Progress* 2021; **20**: 1365–1381.

46. Yan BC, Rabbani G, Lee NLY, Ooi JLS, Lee JN, Huang D, et al. The microbiome of the seagrass *Halophila ovalis*: community structuring from plant parts to regional scales. *Aquatic Microbial Ecology* 2021; **87**: 139–150.

47. Graham OJ, Adamczyk EM, Schenk S, Dawkins P, Burke S, Chei E, et al. Manipulation of the seagrass-associated microbiome reduces disease severity. *Environmental Microbiology* 2024; **26**: e16582.

48. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* 2018; **6**: 42.

49. Gan M, Wu B, Yan G, Li G, Sun L, Lu G, et al. Combined nanopore adaptive sequencing and enzyme-based host depletion efficiently enriched microbial sequences and identified missing respiratory pathogens. *BMC Genomics* 2021; **22**: 732.

50. Heravi FS, Zakrzewski M, Vickery K, Hu H. Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *Journal of Microbiological Methods* 2020; **170**: 105856.

51. Song L, Xie K. Engineering CRISPR/Cas9 to mitigate abundant host contamination for 16S rRNA gene-based amplicon sequencing. *Microbiome* 2020; **8**: 80.

52. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, et al. Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 2012; **488**: 86–90.

53. Zarraonaindia I, Owens SM, Weisenhorn P, West K, Hampton-Marcell J, Lax S, et al. The soil microbiome influences grapevine-associated microbiota. *mBio* 2015; **6**: e02527-14.

54. Oortwijn T, de Fouw J, Petersen JM, van Gils JA. Sulfur in lucinid bivalves inhibits intake rates of a molluscivore shorebird. *Oecologia* 2022; **199**: 69–78.

55. Fouz MF, Appella DH. PNA clamping in nucleic acid amplification protocols to detect single nucleotide mutations related to cancer. *Molecules* 2020; **25**: 786.

56. Kawasaki A, Ryan PR. Peptide Nucleic Acid (PNA) clamps to reduce co-amplification of plant DNA during PCR amplification of 16S rRNA genes from endophytic bacteria. *The Plant Microbiome. Methods in Molecular Biology*. 2021. pp 123–134.

57. Ørum H, Nielsen PE, Egholm M, Berg RH, Buchardt O, Stanley C. Single base pair mutation analysis by PNA directed PCR clamping. *Nucleic Acids Research* 1993; **21**: 5332–5336.

58. Buytaers FE, Saltykova A, Denayer S, Verhaegen B, Vanneste K, Roosens NHC, et al. Towards real-time and affordable strain-level metagenomics-based foodborne outbreak investigations using Oxford nanopore sequencing technologies. *Frontiers in Microbiology* 2021; **12**.

59. Doualeh M, Payne M, Litton E, Raby E, Currie A. Molecular methodologies for improved polymicrobial sepsis diagnosis. *International Journal of Molecular Sciences* 2022; **23**: 4484.

60. Shi Y, Wang G, Lau HC-H, Yu J. Metagenomic sequencing for microbial DNA in human samples: emerging technological advances. *International Journal of Molecular Sciences* 2022; **23**: 2181.

61. Marchukov D, Li J, Juillerat P, Misselwitz B, Yilmaz B. Benchmarking microbial DNA enrichment protocols from human intestinal biopsies. *Frontiers in Genetics* 2023; **14**.

62. Wright ML, Podnar J, Longoria KD, Nguyen TC, Lim S, Garcia S, et al. Comparison of commercial DNA extraction kits for whole metagenome sequencing of human oral, vaginal, and rectal microbiome samples. 2023. bioRxiv. , 2023.02.01.526597.

63. Bloomfield SJ, Zomer AL, O'Grady J, Kay GL, Wain J, Janecko N, et al. Determination and quantification of microbial communities and antimicrobial resistance on food through host DNA-depleted metagenomics. *Food Microbiology* 2023; **110**: 104162.

64. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013; **8**: e61217.

65. Oksanen J, Simpson G, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: community ecology package. 2024.

66. Shetty S, Lahti L. microbiomeutilities: microbiomeutilities: utilities for microbiome analytics. 2022.

67. Barnett DJM, Arts ICW, Penders J. microViz: an R package for microbiome data visualization and statistics. 2021.

68. Arnholt AT, Evans B. BSDA: basic statistics and data analysis. 2022.

69. Bittinger K. usedist: Distance Matrix Utilities. 2024.

70. Smith RJ. phytomosaic/ecole: ecole: school of ecology package. 2021.

71. Shetty SA, Hermes GDA. Microbial diversity analysis using R tools. 2018.

72. Gao C-H, Chen C, Akyol T, Dusa A, Yu G, Cao B, et al. ggVennDiagram: intuitive venn diagram software extended. *iMeta* 2024; **3**: e177.

73. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 2014; **15**: 550.

74. Wickham H. ggplot2: elegant graphics for data analysis. 2016. Springer-Verlag, New Yorker.

75. Auguie B, Antonov A. gridExtra: miscellaneous functions for 'grid' graphics. 2017.

76. Pedersen TL. patchwork: the composer of plots. 2024.

77. Wickham H, Pedersen TL. scales: scale functions for visualization. 2023.

78. Barrett T, Dowle M, Srinivasan A, Gorecki J, Chirico M, Hocking T, et al. data.table: extension of `data.frame`. 2024.

79. Wickham H, François R, Henry L, Müller K, Vaughna D. dplyr: a grammar of data manipulation. 2023.

80. Iannone R, Cheng J, Schloerke B, Hughes E, Lauer A, Seo J, et al. gt: easily create presentation-ready display tables. 2024.

81. Harrell FE. Hmisc: Harrell miscellaneous. 2024.

82. Wickham H. Reshaping data with the reshape package. *Journal of Statistical Software* 2007; **21**: 1–20.

83. Xie Y. knitr: a general-purpose package for dynamic report generation in R. 2024.

84. Wickham H. The split-apply-combine strategy for data analysis. *Journal of Statistical Software* 2011; **40**: 1–29.

85. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. *Journal of Open Source Software* 2019; **4**: 1686.

86. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nature Biotechnology* 2019; **37**: 783–792.

87. Bruggeling CE, Garza DR, Achouiti S, Mes W, Dutilh BE, Boleij A. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. *MicrobiologyOpen* 2021; **10**: e1191.

88. Lewis ZJ, Scott A, Madden C, Vik D, Zayed AA, Smith GJ, et al. Evaluating urine volume and host depletion methods to enable genome-resolved metagenomics of the urobiome. 2024. Research Square.

89. Wu-Woods NJ, Barlow JT, Trigodet F, Shaw DG, Romano AE, Jabri B, et al. Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples. *Nature Methods* 2023; **20**: 1672–1682.

90. QIAamp DNA Microbiome Kit. *QIAGEN*. https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/microbial-dna/qiaamp-dna-microbiome-kit. Accessed 13 Oct 2024.

91. HostZERO Microbial DNA Kit. *Zymo Research International*. https://zymoresearch.eu/products/hostzero-microbial-dna-kit. Accessed 13 Oct 2024.

92. PNA Bio. mPNA & pPNA. *PNA Bio*. https://www.pnabio.com/products/PCR_blocker.htm. Accessed 20 Sep 2024.

93. Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL. Practical innovations for high-throughput amplicon sequencing. *Nature Methods* 2013; **10**: 999–1002.

94. Cregger MA, Veach AM, Yang ZK, Crouch MJ, Vilgalys R, Tuskan GA, et al. The *Populus* holobiont: dissecting the effects of plant niches and genotype on the microbiome. *Microbiome* 2018; **6**: 31.

95. Fitzpatrick CR, Lu-Irving P, Copeland J, Guttman DS, Wang PW, Baltrus DA, et al. Chloroplast sequence variation and the efficacy of peptide nucleic acids for blocking host amplification in plant microbiome studies. *Microbiome* 2018; **6**: 144.

96. NCBI. *Zostera marina* voucher C2544 mitochondrion, complete genome. 2017.

97. NCBI. *Zostera marina* chloroplast, complete genome. 2023.

98. Lee E, Lee K-W, Park Y, Choi A, Kwon KK, Kang H-M. Comparative microbiome analysis of *Artemia* spp. and potential role of microbiota in cyst hatching. *Marine Biotechnology* 2024; **26**: 50–59.

99. Serag AM, Abdel-Sabour MS, El-Hadidi M, Maged M, Magdy M, Ramadan MF, et al. Comparative 16S Metabarcoding of Nile tilapia gut microbiota from the Northern Lakes of Egypt. *Applied Biochemistry and Biotechnology* 2022; **194**: 2168–2182.

100. Tarnecki AM, Wafapoor M, Phillips RN, Rhody NR. Benefits of a *Bacillus* probiotic to larval fish survival and transport stress resistance. *Scientific Reports* 2019; **9**: 4892.

101. Bingöl A, Er A, İpek ZZ, Kayış Ş. Bacterial examination of wild and cultured fish present in the same aquatic ecosystem, and the antibiotic resistance of the isolated bacteria. *Genetics of Aquatic Organisms* 2021; **6**.

102. Su H-Y, Hussain B, Hsu B-M, Lee K-H, Mao Y-C, Chiang L-C, et al. Bacterial community analysis identifies *Klebsiella pneumoniae* as a native symbiotic bacterium in the newborn *Protobothrops mucrosquamatus*. *BMC Microbiology* 2023; **23**: 213.

103. Suppa A. Phenotypic plasticity in the *Daphnia* model organism: gene expression and maternal control in the adaptation to environmental perturbations. 2019. Università degli studi di Ferrara.

104. Shin H, Lee E, Shin J, Ko S-R, Oh H-S, Ahn C-Y, et al. Elucidation of the bacterial communities associated with the harmful microalgae *Alexandrium tamarense* and *Cochlodinium polykrikoides* using nanopore sequencing. *Scientific Reports* 2018; **8**: 5323.

105. Ferguson LV, Dhakal P, Lebenzon JE, Heinrichs DE, Bucking C, Sinclair BJ. Seasonal shifts in the insect gut microbiome are concurrent with changes in cold tolerance and immunity. *Functional Ecology* 2018; **32**: 2357–2368.

106. Donkersley P, Rice A, Graham RI, Wilson K. Gut microbial community supplementation and reduction modulates African armyworm susceptibility to a baculovirus. *FEMS Microbiology Ecology* 2023; **99**: fiac147.

107. Gómez-Govea MA, de Lourdes Ramírez-Ahuja M, Castellanos-López LM, Ruvalcaba-Ortega I, Reyes-Cortes LM, de Jesús Trujillo-Rodríguez G, et al. Bacterial communities associated with *Megalopyge opercularis* (Smith) (Lepidoptera: Megalopygidae): exploring poisonous lepidopterans. *The Florida Entomologist* 2022; **105**: 298–306.

108. Lenka JL. Identification and characterisation of lignin degrading bacteria and enzymes from larval guts of the African palm weevil (*Rhynchophorus Phoenicis*). 2022. University of Salford.

109. Cardoso A, Gómez-Zurita J. Food Resource Sharing of alder leaf beetle specialists (Coleoptera: Chrysomelidae) as potential insect–plant interface for horizontal transmission of endosymbionts. *Environmental Entomology* 2020; **49**: 1402–1414.

110. E Silva B, Matsena Zingoni Z, Koekemoer LL, Dahan-Moss YL. Microbiota identified from preserved *Anopheles*. *Malaria Journal* 2021; **20**: 230.

111. Paulraj S, Bhat R, Rajesh MK, Ramesh SV, Priya UK, Pandian RTP, et al. Data of 16S rRNA gene amplicon-based metagenomic signatures of arecanut rhizosphere soils in Yellow Leaf Disease (YLD) endemic region of India. *Data in Brief* 2021; **38**: 107443.

112. Takano S, Gotoh Y, Hayashi T. "*Candidatus* Mesenet longicola": novel endosymbionts of *Brontispa longissima* that induce cytoplasmic incompatibility. *Microbial Ecology* 2021; **82**: 512–522.

113. Hill JL, Moore A, Williams RTP, Nishimura EO. Hand dissection of *Caenorhabditis elegans* intestines. *Journal of Visualized Experiments* 2022; e64120.

114. Liu C, Zhang J, Shao Z, Xia X, Lyu Y, Xie F, et al. Insights into the *Galaxea fascicularis* microbiome obtained from the microenvironment-based investigation. *Ecological Indicators* 2024; **159**: 111627.

115. León AV-P de, Hoetzinger M, Hensen T, Gupta S, Weston B, Johnsen SM, et al. The salmon microbial genome atlas enables novel insights into bacteria-host interactions via functional mapping. 2023. bioRxiv. , 2023.12.10.570985.

116. Maire J, Collingro A, Tandon K, Jameson VJ, Judd LM, Horn M, et al. Chlamydiae as symbionts of photosynthetic dinoflagellates. *The ISME Journal* 2024; **18**: wrae139.

117. Li Q, Chen Y, Zhang S, Lyu Y, Zou Y, Li J. DNA enrichment methods for microbial symbionts in marine bivalves. *Microorganisms* 2022; **10**: 393.

118. Underwood LH. Habitat value of green-lipped mussel (*Perna canaliculus*) farms for fish in northern Aotearoa New Zealand. 2023. Thesis, University of Auckland.

119. Iqbal MM, Nishimura M, Haider MdN, Sano M, Ijichi M, Kogure K, et al. Diversity and composition of microbial communities in an eelgrass (*Zostera marina*) bed in Tokyo Bay, Japan. *Microbes and Environments* 2021; **36**.

120. Machairas F. “Finding Symbo” diversity and distribution of widespread symbionts of marine lucinid clams in the surrounding environment. 2022. University of Ioannina.

121. Chistoserdova L, Kalyuzhnaya MG. Current trends in methylotrophy. *Trends in Microbiology* 2018; **26**: 703–714.

122. Abanda-Nkpwatt D, Müsch M, Tschiersch J, Boettner M, Schwab W. Molecular interaction between Methylobacterium extorquens and seedlings: growth promotion, methanol consumption, and localization of the methanol emission site. *Journal of Experimental Botany* 2006; **57**: 4025–4032.

123. Crump BC, Wojahn JM, Tomas F, Mueller RS. Metatranscriptomics and amplicon sequencing reveal mutualisms in seagrass microbiomes. *Frontiers in Microbiology* 2018; **9**.

124. Gosink JJ, Woese CR, Staley JT. *Polaribacter* gen. nov., with three new species, *P. irgensii* sp. nov., *P. franzmannii* sp. nov. and *P. filamentus* sp. nov., gas vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of ‘*Flectobacillus glomeratus*’ as *Polaribacter glomeratus* comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 1998; **48**: 223–235.

125. Park S, Park JM, Jung YT, Lee KH, Yoon JH. *Polaribacter undariae* sp. nov., isolated from a brown alga reservoir. *International Journal of Systematic and Evolutionary Microbiology* 2015; **65**: 1679-1685.

126. Xu W, Chen XY, Wei XT, Lu DC, Du ZJ. *Polaribacter aquimarinus* sp. nov., isolated from the surface of a marine red alga. *Antonie van Leeuwenhoek* 2020; **113**: 407-415.

127. Fukui Y, Abe M, Kobayashi M, Saito H, Oikawa H, Yano Y, et al. *Polaribacter porphyrae* sp. nov., isolated from the red alga *Porphyra yezoensis*, and emended descriptions of the genus *Polaribacter* and two *Polaribacter* species. *International Journal of Systematic and Evolutionary Microbiology*; **63**: 1665-1672.

128. Nedashkovskaya OI, Kukhlevskiy AD, Zhukova NV. *Polaribacter reichenbachii* sp. nov.: a new marine bacterium associated with the green alga *Ulva fenestrata*. *Current Microbiology* 2013; **66**: 16-21.

129. Nedashkovskaya OI, Kim SG, Balabanova LA, Zhukova NV, Bakunina IY, Mikhailov VV. *Polaribacter staleyi* sp. nov., a polysaccharide-degrading marine bacterium isolated from the red alga *Ahnfeltia tobuchiensis*. *International Journal of Systematic and Evolutionary Microbiology* 2018; **68**: 623-629.

130. Lozada M, Diéguez MC, García PE, Dionisi HM. Microbial communities associated with kelp detritus in temperate and subantarctic intertidal sediments. *Science of The Total Environment* 2023; **857**: 159392.

131. Zhang L, Li X, Zhang X, Li Y, Wang L. Bacterial alginate metabolism: an important pathway for bioconversion of brown algae. *Biotechnology for Biofuels* 2021; **14**: 158

132. Sun H, Wang T, Liu S, Tang X, Sun J, Liu X, et al. Novel insights into the rhizosphere and seawater microbiome of *Zostera marina* in diverse mariculture zones. *Microbiome* 2024; **12**: 27.

133. Minich JJ, Morris MM, Brown M, Doane M, Edwards MS, Michael TP, et al. Elevated temperature drives kelp microbiome dysbiosis, while elevated carbon dioxide induces water microbiome disruption. *PLoS ONE* 2018; **13**: e0192772.

134. Avcı B, Krüger K, Fuchs BM, Teeling H, Amann RI. Polysaccharide niche partitioning of distinct *Polaribacter* clades during North Sea spring algal blooms. *The ISME Journal* 2020; **14**: 1369–1383.

135. Crump BC, Wojahn JM, Tomas F, Mueller RS. Frontiers | Metatranscriptomics and Amplicon Sequencing Reveal Mutualisms in Seagrass Microbiomes. *Frontiers in Microbiology* 2018; **9**.

136. Lee K, Lee HK, Choi T-H, Cho J-C. *Robiginitomaculum antarcticum* gen. nov., sp. nov., a member of the family Hyphomonadaceae, from Antarctic seawater. *International Journal of Systematic and Evolutionary Microbiology* 2007; **57**: 2595–2599.

137. Ramírez-Puebla ST, Weigel BL, Jack L, Schlundt C, Pfister CA, Mark Welch JL. Spatial organization of the kelp microbiome at micron scales. *Microbiome* 2022; **10**: 52.

138. Lu D-C, Wang F-Q, Amann RI, Teeling H, Du Z-J. Epiphytic common core bacteria in the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae. *Microbiome* 2023; **11**: 126.

139. Weigel BL, Pfister CA. Successional dynamics and seascape-level patterns of microbial communities on the canopy-forming kelps *Nereocystis luetkeana* and *Macrocystis pyrifera*. *Frontiers in Microbiology* 2019; **10**.

140. Sorokin D. *Sulfitobacter pontiacus* gen. nov., sp. nov. - A new heterotrophic bacterium from the Black Sea, specialized on sulfite oxidation. *Microbiology* 1995; **64**: 295–305.

141. Prabagaran SR, Manorama R, Delille D, Shivaji S. Predominance of *Roseobacter*, *Sulfitobacter*, *Glaciecola* and *Psychrobacter* in seawater collected off Ushuaia, Argentina, Sub-Antarctica. *FEMS Microbiology Ecology* 2007; **59**: 342–355.

142. Yang Q, Ge Y, Iqbal NM, Yang X, Zhang X. *Sulfitobacter alexandrii* sp. nov., a new microalgae growth-promoting bacterium with exopolysaccharides bioflocculanting potential isolated from marine phycosphere. *Antonie van Leeuwenhoek* 2021; **114**: 1091–1106.

143. Beiralas R, Ozer N, Segev E. Abundant Sulfitobacter marine bacteria protect *Emiliania huxleyi* algae from pathogenic bacteria. *ISME Communications* 2023; **3**: 1–10.

144. Barak-Gavish N, Frada MJ, Ku C, Lee PA, DiTullio GR, Malitsky S, et al. Bacterial virulence against an oceanic bloom-forming phytoplankter is mediated by algal DMSP. *Science Advances* 2018; **4**: eaau5716.

145. Zhang F, Fan Y, Zhang D, Chen S, Bai X, Ma X, et al. Effect and mechanism of the algicidal bacterium *Sulfitobacter porphyrae* ZFX1 on the mitigation of harmful algal blooms caused by *Prorocentrum donghaiense*. *Environmental Pollution* 2020; **263**: 114475.

146. Espinosa E, Marco-Noales E, Gómez D, Lucas-Elío P, Ordax M, Garcías-Bonet N, et al. Taxonomic study of *Marinomonas strains* isolated from the seagrass *Posidonia oceanica*, with descriptions of *Marinomonas balearica* sp. nov. and *Marinomonas pollencensis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 2010; **60**: 93–98.

147. Ranson HJ, LaPorte J, Spinard E, Chistoserdov AY, Gomez-Chiarri M, Nelson DR, et al. Draft genome sequence of the putative marine pathogen *Aquimarina* sp. Strain I32.4. *Genome Announcements* 2018; **6**: e00313-18.

148. Ling S-K, Xia J, Liu Y, Chen G-J, Du Z-J. *Agarilytica rhodophyticola* gen. nov., sp. nov., isolated from *Gracilaria blodgettii*. *International Journal of Systematic and Evolutionary Microbiology* 2017; **67**: 3778–3783.

149. Yoon J, Adachi K, Kasai H. Isolation and characterization of a novel marine Bacteroidetes as *Algitalea ulvae* gen. nov., sp. nov., isolated from the green alga *Ulva pertusa*. *Antonie Van Leeuwenhoek* 2015; **108**: 505–513.

150. Nedashkovskaya OI, Vancanneyt M, Christiaens L, Kalinovskaya NI, Mikhailov VV, Swings J. *Aquimarina intermedia* sp. nov., reclassification of *Stanierella latercula* (Lewin 1969) as *Aquimarina latercula* comb. nov. and *Gaetbulimicrobium brevivitae* Yoon et al. 2006 as *Aquimarina brevivitae* comb. nov. and emended description of the genus *Aquimarina*. *International Journal of Systematic and Evolutionary Microbiology* 2006; **56**: 2037–2041.

151. Nedashkovskaya OI, Kim SB, Lysenko AM, Frolova GM, Mikhailov VV, Lee KH, et al. Description of *Aquimarina muelleri* gen. nov., sp. nov., and proposal of the reclassification of [*Cytophaga*] *latercula* Lewin 1969 as *Stanierella latercula* gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 2005; **55**: 225–229.

152. Romanenko LA, Tanaka N, Frolova GM, Mikhailov VV. *Arenicella xantha* gen. nov., sp. nov., a gammaproteobacterium isolated from a marine sandy sediment. *International Journal of Systematic and Evolutionary Microbiology* 2010; **60**: 1832–1836.

153. Schlesner H, Rensmann C, Tindall BJ, Gade D, Rabus R, Pfeiffer S, et al. Taxonomic heterogeneity within the Planctomycetales as derived by DNA–DNA hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov. and emended description of the genus *Pirellula*. *International Journal of Systematic and Evolutionary Microbiology* 2004; **54**: 1567–1580.

154. Ivanova EP, Webb H, Christen R, Zhukova NV, Kurilenko VV, Kalinovskaya NI, et al. *Celeribacter neptunius* gen. nov., sp. nov., a new member of the class Alphaproteobacteria. *International Journal of Systematic and Evolutionary Microbiology* 2010; **60**: 1620–1625.

155. Lee K, Lee HK, Choi T-H, Kim K-M, Cho J-C. Granulosicoccaceae fam. nov., to include *Granulosicoccus antarcticus* gen. nov., sp. nov., a non-phototrophic, obligately aerobic chemoheterotroph in the order Chromatiales, isolated from Antarctic seawater. *Journal of Microbiology and Biotechnology* 2007; **17**: 1483–1490.

156. Anagnostidis K, Komárek J. Modern approach to the classification system of cyanophytes. 3 - Oscillatoriales. *Algological Studies* 1988; 327–472.

157. Lasa A, Pichon P, Diéguez AL, Romalde JL. *Marinomonas gallaica* sp. nov. and *Marinomonas atlantica* sp. nov., isolated from reared clams (*Ruditapes decussatus*). *International Journal of Systematic and Evolutionary Microbiology* 2016; **66**: 3183–3188.

158. Romanenko LA, Uchino M, Mikhailov VV, Zhukova NV, Uchimura T. *Marinomonas primoryensis* sp. nov., a novel psychrophile isolated from coastal sea-ice in the Sea of Japan. *International Journal of Systematic and Evolutionary Microbiology* 2003; **53**: 829–832.

159. Marinomonas aquamarina sp. nov., isolated from oysters and seawater - ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0723202004000773?via%3Dihub. Accessed 10 Jul 2024.

160. Seo H-S, Kwon KK, Yang S-H, Lee H-S, Bae SS, Lee J-H, et al. *Marinoscillum* gen. nov., a member of the family 'Flexibacteraceae', with *Marinoscillum pacificum* sp. nov. from a marine sponge and *Marinoscillum furvescens* nom. rev., comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 2009; **59**: 1204–1208.

161. Allaby M. LPP group. *Oxford Dictionary of Plant Sciences* . 2019. Oxford University Press, Oxford.

162. Nakamura Y, Takahashi J, Sakurai A, Inaba Y, Suzuki E, Nihei S, et al. Some cyanobacteria synthesize semi-amylopectin type α-polyglucans instead of glycogen. *Plant and Cell Physiology* 2005; **46**: 539–545.

163. Kalyuzhnaya MG, Bowerman S, Lara JC, Lidstrom ME, Chistoserdova L. *Methylotenera mobilis* gen. nov., sp. nov., an obligately methylamine-utilizing bacterium within the family Methylophilaceae. *International Journal of Systematic and Evolutionary Microbiology* 2006; **56**: 2819–2823.

164. Golyshin PN, Chernikova TN, Abraham W-R, Lünsdorf H, Timmis KN, Yakimov MM. Oleiphilaceae fam. nov., to include *Oleiphilus messinensis* gen. nov., sp. nov., a novel marine bacterium that obligately utilizes hydrocarbons. *International Journal of Systematic and Evolutionary Microbiology* 2002; **52**: 901–911.

165. Yoon J, Matsuo Y, Adachi K, Nozawa M, Matsuda S, Kasai H, et al. Description of *Persicirhabdus sediminis* gen. nov., sp. nov., *Roseibacillus ishigakijimensis* gen. nov., sp. nov., *Roseibacillus ponti* sp. nov., *Roseibacillus persicicus* sp. nov., *Luteolibacter pohnpeiensis* gen. nov., sp. nov. and *Luteolibacter algae* sp. nov., six marine members of the phylum 'Verrucomicrobia', and emended descriptions of the class Verrucomicrobiae, the order Verrucomicrobiales and the family Verrucomicrobiaceae. *International Journal of Systematic and Evolutionary Microbiology* 2008; **58**: 998–1007.

166. Davydov D, Vilnet A. Review of the cyanobacterial genus *Phormidesmis* (Leptolyngbyaceae) with the description of *Apatinema* gen. nov. *Diversity* 2022; **14**: 731

167. Yoon J, Matsuo Y, Kasai H, Yokota A. *Portibacter lacus* gen. nov., sp.nov., a new member of the family Saprospiraceae isolated from a saline lake. *The Journal of General and Applied Microbiology* 2012; **58**: 191–197.

168. Yoon J, Katsuta A, Kasai H. *Rubidimonas crustatorum* gen. nov., sp. nov., a novel member of the family Saprospiraceae isolated from a marine crustacean. *Antonie van Leeuwenhoek* 2012; **101**: 461–467.

169. Scheuermayer M, Gulder TAM, Bringmann G, Hentschel U. *Rubritalea marina* gen. nov., sp. nov., a marine representative of the phylum 'Verrucomicrobia', isolated from a sponge (Porifera). *International Journal of Systematic and Evolutionary Microbiology* 2006; **56**: 2119–2124.

170. Gomont M. Monographie des Oscillariées (Nostocacées homocystées). 1892; **7**: 263–368.

171. Wirth JS, Whitman WB. Phylogenomic analyses of a clade within the roseobacter group suggest taxonomic reassignments of species of the genera *Aestuariivita*, *Citreicella*, *Loktanella*, *Nautella*, *Pelagibaca*, *Ruegeria*, *Thalassobius*, *Thiobacimonas* and *Tropicibacter*, and the proposal of six novel genera. *International Journal of Systematic and Evolutionary Microbiology* 2018; **68**: 2393–2411.

172. Lewin RA. *Flexithrix dorotheae* gen. et sp. nov. (Flexibacterales); and suggestions for reclassifying sheathed bacteria. *Canadian Journal of Microbiology* 1970; **16**: 511–515.

173. Bowman JP, McCammon SA, Brown JL, McMeekin TA. *Glaciecola punicea* gen. nov., sp. nov. and *Glaciecola pallidula* gen. nov., sp. nov.: psychrophilic bacteria from Antarctic sea-ice habitats. *International Journal of Systematic and Evolutionary Microbiology* 1998; **48**: 1213–1222.

174. Tasdemir D, Scarpato S, Utermann-Thüsing C, Jensen T, Blümel M, Wenzel-Storjohann A, et al. Epiphytic and endophytic microbiome of the seagrass *Zostera marina*: do they contribute to pathogen reduction in seawater? *Science of The Total Environment* 2024; **908**: 168422.

175. McCauley EP, Haltli B, Kerr RG. Description of *Pseudobacteriovorax antillogorgiicola* gen. nov., sp. nov., a bacterium isolated from the gorgonian octocoral *Antillogorgia elisabethae*, belonging to the family Pseudobacteriovoracaceae fam. nov., within the order Bdellovibrionales. *International Journal of Systematic and Evolutionary Microbiology* 2015; **65**: 522-530.

176. Nedashkovskaya OI, Kim SB, Shin DS, Beleneva IA, Mikhailov VV. *Fulvivirga kasyanovii* gen. nov., sp. nov., a novel member of the phylum Bacteroidetes isolated from seawater in a mussel farm. *International Journal of Systematic and Evolutionary Microbiology* 2007; **57**: 1046–1049.

177. Nguyen TTH, Vuong TQ, Han HL, Li Z, Lee YJ, Ko J, et al. Three marine species of the genus *Fulvivirga*, rich sources of carbohydrate-active enzymes degrading alginate, chitin, laminarin, starch, and xylan. *Scientific Reports* 2023; **13**: 6301

178. Thrash JC, Coates JD. Phylum XVII. Acidobacteria phyl. nov. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, et al. (eds). *Bergey's Manual® of Systematic Bacteriology: Volume Four The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes*. 2010. Springer, New York, NY, pp 725–735.

179. Hauck F. Die Meeresalgen Deutschlands und Österreichs. *Dr. L. Rabenhorst's Kryptogamen-Flora von Deutschland, Oesterreich und der Schweiz*. 1885. Leipzig, pp 513–575.

180. Shalygin S, Kavulic KJ, Pietrasiak N, Bohunická M, Vaccarino MA, Chesarino NM, et al. Neotypification of *Pleurocapsa fuliginosa* and epitypification of *P. minor* (Pleurocapsales): resolving a polyphyletic cyanobacterial genus. *Phytotaxa* 2019; **392**: 245.

181. Shivaji S, Reddy GS. Phylogenetic analyses of the genus *Glaciecola*: emended description of the genus *Glaciecola*, transfer of *Glaciecola mesophila*, *G. agarilytica*, *G. aquimarina*, *G. arctica*, *G. chathamensis*, *G. polaris* and *G. psychrophila* to the genus *Paraglaciecola* gen. nov. as *Paraglaciecola mesophila* comb. nov., *P. agarilytica* comb. nov., *P. aquimarina* comb. nov., *P. arctica* comb. nov., *P. chathamensis* comb. nov., *P. polaris* comb. nov. and *P. psychrophila* comb. nov., and description of *Paraglaciecola oceanifecundans* sp. nov., isolated from the Southern Ocean. *International Journal of Systematic and Evolutionary Microbiology* 2014; **64**: 3264–3275.

182. Yoon J, Matsuo Y, Matsuda S, Adachi K, Kasai H, Yokota A. *Cerasicoccus arenae* gen. nov., sp. nov., a carotenoid-producing marine representative of the family Puniceicoccaceae within the phylum 'Verrucomicrobia', isolated from marine sand. *International Journal of Systematic and Evolutionary Microbiology* 2007; **57**: 2067–2072.

183. Jeon CO, Lim J-M, Park D-J, Kim C-J. Salinimonas chungwhensis gen. nov., sp. nov., a moderately halophilic bacterium from a solar saltern in Korea. *International Journal of Systematic and Evolutionary Microbiology* 2005; **55**: 239–243.

184. Cui Z, Li Y, Jing X, Luan X, Liu N, Liu J, et al. Cycloalkane degradation by an uncultivated novel genus of Gammaproteobacteria derived from China's marginal seas. *Journal of Hazardous Materials* 2024; **469**: 133904.

185. Jean WD, Hsu CY, Huang S-P, Chen J-S, Lin S, Su M-H, et al. Reclassification of [*Glaciecola*] *lipolytica* and [*Aestuariibacter*] *litoralis* in *Aliiglaciecola* gen. nov., as *Aliiglaciecola lipolytica* comb. nov. and *Aliiglaciecola litoralis* comb. nov., respectively. *International Journal of Systematic and Evolutionary Microbiology* 2013; **63**: 2859–2864.

186. Khan ST, Nakagawa Y, Harayama S. *Sediminibacter furfurosus* gen. nov., sp. nov. and *Gilvibacter sediminis* gen. nov., sp. nov., novel members of the family Flavobacteriaceae. *International Journal of Systematic and Evolutionary Microbiology* 2007; **57**: 265–269.

187. Yoon J, Yasumoto-Hirose M, Matsuo Y, Nozawa M, Matsuda S, Kasai H, et al. *Pelagicoccus mobilis* gen. nov., sp. nov., *Pelagicoccus albus* sp. nov. and *Pelagicoccus litoralis* sp. nov., three novel members of subdivision 4 within the phylum 'Verrucomicrobia', isolated from seawater by in situ cultivation. *International Journal of Systematic and Evolutionary Microbiology* 2007; **57**: 1377-1385.

188. Sly LI, Taghavit M, Fegan M. Phylogenetic heterogeneity within the genus *Herpetosiphon*: transfer of the marine species *Herpetosiphon cohaerens*, *Herpetosiphon nigricans* and *Herpetosiphon persicus* to the genus *Lewinella* gen. nov. in the *Flexibacter-Bacteroides-Cytophaga* phylum. *International Journal of Systematic and Evolutionary Microbiology* 1998; **48**: 731–737.

189. Bowman JP, Nichols CM, Gibson JAE. *Algoriphagus ratkowskyi* gen. nov., sp. nov., Brumimicrobium glaciale gen. nov., sp. nov., *Cryomorpha ignava* gen. nov., sp. nov. and *Crocinitomix catalasitica* gen. nov., sp. nov., novel flavobacteria isolated from various polar habitats. *International Journal of Systematic and Evolutionary Microbiology* 2003; **53**: 1343-1355.

190. Kurahashi M, Yokota A. *Endozoicomonas elysicola* gen. nov., sp. nov., a *γ*-proteobacterium isolated from the sea slug *Elysia ornata*. *Systematic and Applied Microbiology* 2007; **30**: 202–206.

191. Neave MJ, Michell CT, Apprill A, Voolstra CR. *Endozoicomonas* genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Scientific Reports* 2017; **7**: 40579.

192. Maire J, Tandon K, Collingro A, van de Meene A, Damjanovic K, Gotze CR, et al. Colocalization and potential interactions of *Endozoicomonas* and chlamydiae in microbial aggregates of the coral *Pocillopora acuta*. *Science Advances* 2023; **9**: eadg0773

193. Albuquerque L, Rainey FA, Nobre MF, da Costa MS. *Schleiferia thermophila* gen. nov., sp. nov., a slightly thermophilic bacterium of the phylum 'Bacteroidetes' and the proposal of Schleiferiaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology* 2011; **61**: 2450–2455.

194. Barbeyron T, L'Haridon S, Corre E, Kloareg B, Potin P. *Zobellia galactanovorans* gen. nov., sp. nov., a marine species of Flavobacteriaceae isolated from a red alga, and classification of [*Cytophaga*] *uliginosa* (ZoBell and Upham 1944) Reichenbach 1989 as *Zobellia uliginosa* gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 2001; **51**: 985–997.

195. Romanenko LA, Tanaka N, Frolova GM, Mikhailov VV. *Marinicella litoralis* gen. nov., sp. nov., a gammaproteobacterium isolated from coastal seawater. *International Journal of Systematic and Evolutionary Microbiology* 2010; **60**: 1613–1619.

196. Deming JW, Somers LK, Straube WL, Swartz DG, Macdonell MT. Isolation of an Obligately Barophilic Bacterium and Description of a New Genus, *Colwellia* gen. nov. *Systematic and Applied Microbiology* 1988; **10**: 152–160.

197. Deming JW. Exceeding expectations out in the cold with *Colwellia*. *Nature Microbiology* 2024; **9**: 1159–1160.

198. Urios L, Intertaglia L, Lesongeur F, Lebaron P. *Eionea nigra* gen. nov., sp. nov., a gammaproteobacterium from the Mediterranean Sea. *International Journal of Systematic and Evolutionary Microbiology* 2011; **61**: 1677–1681.

199. Nedashkovskaya OI, Kim SB, Han SK, Rhee MS, Lysenko AM, Falsen E, et al. *Ulvibacter litoralis* gen. nov., sp. nov., a novel member of the family Flavobacteriaceae isolated from the green alga *Ulva fenestrata*. *International Journal of Systematic and Evolutionary Microbiology* 2004; **54**: 119–123.

200. Park S, Lee J-S, Lee K-C, Yoon J-H. *Lentilitoribacter donghaensis* gen. nov., sp. nov., a slowly-growing alphaproteobacterium isolated from coastal seawater. *Antonie van Leeuwenhoek* 2013; **103**: 457–464.

201. Park S, Park J-M, Lee K-C, Bae KS, Yoon J-H. *Boseongicola aestuarii* gen. nov., sp. nov., isolated from a tidal flat sediment. *International Journal of Systematic and Evolutionary Microbiology* 2014; **64**: 2618–2624.

202. Nichols CM, Bowman JP, Guezennec J. *Olleya marilimosa* gen. nov., sp. nov., an exopolysaccharide-producing marine bacterium from the family Flavobacteriaceae, isolated from the Southern Ocean. *International Journal of Systematic and Evolutionary Microbiology* 2005; **55**: 1557–1561.

203. Bondoso J, Albuquerque L, Lobo-da-Cunha A, da Costa MS, Harder J, Lage OM. *Rhodopirellula lusitana* sp. nov. and *Rhodopirellula rubra* sp. nov., isolated from the surface of macroalgae. *Systematic and Applied Microbiology* 2014; **37**: 157–164.

204. Sreya P, Suresh G, Rai A, Ria B, Vighnesh L, Agre VC, et al. Revisiting the taxonomy of the genus *Rhodopirellula* with the proposal for reclassification of the genus to *Rhodopirellula* sensu stricto, *Aporhodopirellula* gen. nov., *Allorhodopirellula* gen. nov. and *Neorhodopirellula* gen. nov. *Antonie van Leeuwenhoek* 2023; **116**: 243–264.

205. Gosink JJ, Herwig RP, Staley JT. *Octadecabacter arcticus* gen. nov., sp. nov., and *O. antarcticus*, sp. nov., nonpigmented, psychrophilic gas vacuolate bacteria from polar sea ice and water. *Systematic and Applied Microbiology* 1997; **20**: 356–365.

206. Jin MS, Kim KH, Baek JH, Kim JM, Jeon CO. *Octadecabacter algicola* sp. nov. and *Octadecabacter dasysiphoniae* sp. nov., isolated from a marine red alga and emended description of the genus *Octadecabacter*. *International Journal of Systematic and Evolutionary Microbiology* 2023; **73**: 005664.

207. Jung Y-T, Park S, Lee J-S, Oh T-K, Yoon J-H. *Pseudahrensia aquimaris* gen. nov., sp. nov., isolated from seawater. *International Journal of Systematic and Evolutionary Microbiology* 2012; **62**: 2056–2061.

208. Jung JY, Kim JM, Jin HM, Kim SY, Park W, Jeon CO. *Litorimonas taeanensis* gen. nov., sp. nov., isolated from a sandy beach. *International Journal of Systematic and Evolutionary Microbiology* 2011; **61**: 1534–1538.

209. Nedashkovskaya OI, Kukhlevskiy AD, Zhukova NV, Kim S-J, Rhee S-K. *Litorimonas cladophorae* sp. nov., a new alphaproteobacterium isolated from the Pacific green alga *Cladophora stimpsoni*, and emended descriptions of the genus *Litorimonas* and *Litorimonas taeaensis*. *Antonie van Leeuwenhoek* 2013; **103**: 1263–1269.

210. Kim Y-G, Hwang CY, Cho BC. *Pelagicola litoralis* gen. nov., sp. nov., isolated from coastal water in Korea. *International Journal of Systematic and Evolutionary Microbiology* 2008; **58**: 2102–2104.

211. Pérez-Cataluña A, Salas-Massó N, Diéguez AL, Balboa S, Lema A, Romalde JL, et al. Frontiers | Revisiting the taxonomy of the genus *Arcobacter*: getting order from the chaos. *Frontiers in Microbiology* 2018; **9**.

212. Teramoto M, Nishijima M. *Amylibacter marinus* gen. nov., sp. nov., isolated from surface seawater. *International Journal of Systematic and Evolutionary Microbiology* 2014; **64**: 4016–4020.

213. Kim Y-O, Park S, Nam B-H, Kang S-J, Hur Y-B, Kim D-G, et al. Description of *Litoreibacter meonggei* sp. nov., isolated from the sea squirt *Halocynthia roretzi*, reclassification of *Thalassobacter arenae* as *Litoreibacter arenae* comb. nov. and emended description of the genus *Litoreibacter* Romanenko et al. 2011. *International Journal of Systematic and Evolutionary Microbiology* 2012; **62**: 1825–1831.

214. Romanenko LA, Tanaka N, Frolova GM, Svetashev VI, Mikhailov VV. *Litoreibacter albidus* gen. nov., sp. nov. and *Litoreibacter janthinus* sp. nov., members of the class Alphaproteobacteria isolated from the seashore. *International Journal of Systematic and Evolutionary Microbiology* 2011; **61**: 148–154.

215. Hahnke S, Tindall BJ, Schumann P, Sperling M, Brinkhoff T, Simon M. *Planktotalea frisia* gen. nov., sp. nov., isolated from the southern North Sea. *International Journal of Systematic and Evolutionary Microbiology* 2012; **62**: 1619–1624.

216. Song J, Choi A, Im M, Joung Y, Yoshizawa S, Cho J-C, et al. *Aurantivirga profunda* gen. nov., sp. nov., isolated from deep-seawater, a novel member of the family Flavobacteriaceae. *International Journal of Systematic and Evolutionary Microbiology* 2015; **65**: 4850–4856.

217. Wiese J, Saha M, Wenzel-Storjohann A, Weinberger F, Schmaljohann R, Imhoff JF. *Vicingus serpentipes* gen. nov., sp. nov., a new member of the Flavobacteriales from the North Sea. *International Journal of Systematic and Evolutionary Microbiology* 2018; **68**: 333–340.

Appendix 1 – Optimised protocols used for the final Host depletion and DNA extraction

Protocol for the DNeasy PowerSoil Pro Kit

1. Add 800 µl of Buffer CD1 to the tube with the ground up sample
2. Vortex briefly to mix
3. Decant the mixture into the PowerBead Pro Tube
4. Briefly vortex again to mix with the beads
5. Secure the PowerBead Pro Tube horizontally on a Vortex Adapter for 1.5–2 ml tubes on a Vortex Genie 2
6. Vortex at the maximum speed for 20 min
7. Centrifuge the PowerBead Pro Tube at 15,000 g for 1 min
8. Transfer the supernatant to a new, clean 2 ml microcentrifuge tube
9. Add 200 µl of Solution CD2 to the tube
10. Vortex the tube for 5 sec
11. Let the tube incubate at room temperature for 5 min
12. Centrifuge the tube at 15,000 g for 1 min
13. Transfer 700 µl of the supernatant to a new, clean 2 ml microcentrifuge tube
14. Add 600 µl of Solution CD3 to the new tube
15. Vortex for 5 sec
16. Let the tube incubate at room temperature for 5 min
17. Load 650 µl of the lysate onto an MB Spin Column
18. Centrifuge the Spin Column at 15,000 g for 1 min
19. Discard the flow-through
20. Load the remaining lysate from step 16 onto the same MB Spin Column
21. Centrifuge the Spin Column at 15,000 g for 1 min
22. Discard the flow-through
23. Place the MB Spin Column on a new, clean 2 ml Collection Tube
24. Add 500 µl of Solution EA on the Spin Column
25. Centrifuge the Spin Column at 15,000 g for 1 min
26. Discard the flow-through
27. Add 500 µl of Solution C5 on the Spin Column
28. Centrifuge the Spin Column at 15,000 g for 1 min
29. Discard the flow-through
30. Place the MB Spin Column on a new, clean 2 ml Collection Tube
31. Centrifuge the Spin Column at 16,000 g for 2 min
32. Discard the flow-through
33. Place the MB Spin Column on a new, clean 2 ml Elution Tube
34. Add 50µl of Solution C6 to the centre of the membrane of the MB Spin Column
35. Let the tube incubate at room temperature for 5 min
36. Centrifuge the Spin Column at 15,000 g for 1 min
37. Add the flow-through again onto the centre of the membrane
38. Let the tube incubate at room temperature for 5 min
39. Centrifuge the Spin Column at 15,000 g for 1 min
40. Store the extracted DNA at -20°C

Protocol for the QIAamp DNA Microbiome Kit

1. Add 1 ml of PBS to ground up sample
2. Resuspend the sample by pipetting up and down
3. Vortex briefly to mix
4. Place a 40 µm cell strainer on a 50 ml Falcon tube
5. Transfer the mixture onto the cell strainer
6. Briefly centrifuge the Falcon tube to get the sample through the strainer
7. Transfer the liquid into new, clean 2 ml microcentrifuge tube
8. Centrifuge the tube 10 min at 10,000 g
9. Discard the supernatant
10. Add 190 µl of Buffer RDD to the tube
11. Resuspend the pellet by pipetting up and down 10 times
12. Add 2.5 µl Benzonase to the tube
13. Incubate the tube at 37°C for 30 min, shaking at 600 rpm
14. Add 20 µl Proteinase K to the tube
15. Incubate the tube at 56°C for 30 min, shaking at 600 rpm
16. Briefly spin the tube at very low speed to remove condensation
17. Add 100 µl of Reagent DX to 15 ml of Buffer ATL
18. Add 200µl of Buffer ATL with Reagent DX to the tube
19. Transfer the sample into a Pathogen Lysis Tube L
20. Process the tube on a FastPrep 24 device three times for 45 sec at 6.5 m/s, with breaks of 5 min each in between, keep the tubes on ice during the breaks
21. Incubate the tube 5 min at 95°C
22. Centrifuge the tube at 10,000 g for 1 min to reduce the amount of foam
23. Transfer the supernatant into a new, clean 2 ml microcentrifuge tube
24. Add 40 µl of Proteinase K to the tube
25. Mix by vortexing
26. Incubate the tube at 56°C for 30 min, shaking at 600 rpm
27. Add 200 µl of Buffer APL2
28. Mix by pulse vortexing for 30 sec
29. Incubate the tube at 70°C for 10 min
30. Briefly centrifuge the tube at a very low speed to remove condensation
31. Add 200 µl of 100% ethanol to the tube
32. Mix by pulse vortexing for 30 sec
33. Transfer 700 µl of this mixture to a QIAamp UCP Mini Column on a new, clean Collection Tube
34. Centrifuge the tube at 6,000 g for 1 min
35. Discard the flow-through
36. Load the remaining lysate from step 32 onto the same QIAamp UCP Mini Column
37. Centrifuge the tube at 6,000 g for 1 min
38. Discard the flow-through
39. Place the QIAamp UCP Mini Column on a new, clean 2 ml Collection Tube
40. Add 500 µl of Buffer AW1 to the column
41. Centrifuge the tube at 6,000 g for 1 min
42. Place the QIAamp UCP Mini Column on a new, clean 2 ml Collection Tube
43. Add 500 µl of Buffer AW2 to the column

44. Centrifuge the tube at 20,000 g for 3 min
45. Place the QIAamp UCP Mini Column on a new, clean 2 ml Collection Tube
46. Centrifuge the tube at 20,000 g for 1 min
47. Place the QIAamp UCP Mini Column on a new, clean Elution Tube
48. Add 30 µl Buffer AVE to the centre of the membrane of the QIAamp UCP Mini Column
49. Let the tube incubate at room temperature for 5 min
50. Centrifuge the Spin Column at 6,000 g for 1 min
51. Add the flow-through again onto the centre of the membrane
52. Let the tube incubate at room temperature for 5 min
53. Centrifuge the Spin Column at 6,000 g for 1 min
54. Store the extracted DNA at -20°C

Protocol for the HostZERO Microbial DNA Kit

1. Add 1 ml of PBS to ground up sample
2. Resuspend the sample by pipetting up and down
3. Vortex briefly to mix
4. Place a 40 µm cell strainer on a 50 ml Falcon tube
5. Transfer the mixture onto the cell strainer
6. Briefly centrifuge the Falcon tube to get the sample through the strainer
7. Transfer the liquid into new, clean 2 ml microcentrifuge tube
8. Centrifuge the tube 5 min at 10,000 g
9. Discard the supernatant
10. Add 100 µl of Microbial Selection Buffer to the pellet
11. Resuspend the pellet by pipetting up and down 10 times
12. Add 1 µl of Microbial Selection Enzyme
13. Vortex briefly to mix
14. Incubate the tube at 37°C for 30 min
15. Add 20 µl of Proteinase K
16. Vortex for 10 sec
17. Incubate the tube at 55°C for 10 min
18. Add 100 µl of a 2x concentrate of DNA/RNA Shield
19. Vortex for 10 sec
20. Incubate the tube at room temperature for 5 min
21. Transfer the whole content to a new, clean ZR BashingBead™ Lysis Tube (0.1 & 0.5 mm)
22. Add 750 µl of ZymoBIOMICS™ Lysis Solution
23. Process the tube on a FastPrep 24 device three times for 1 min at 6.5 m/s, with breaks of 5 min each in between, keep the tubes on ice during the breaks
24. Centrifuge the tube 1 min at 10,000 g
25. Transfer the supernatant to a new, clean Collection Tube
26. Add 1.2 ml of DNA Binding Buffer to the Collection Tube
27. Mix by pipetting up and down 10 times
28. Transfer 650µl of this mixture onto a Zymo-Spin™ IC-Z Column4 in a new, clean Collection Tube
29. Centrifuge the Spin Column at 10,000 g for 1 min
30. Discard the flow-through
31. Load the remaining mixture from step 27 onto the same Zymo-Spin™ IC-Z Column4

32. Centrifuge the Spin Column at 10,000 g for 1 min
33. Discard the flow-through
34. Transfer the column to a new, clean Collection Tube
35. Add 400 µl of DNA Wash Buffer 1 to the column
36. Centrifuge the Spin Column at 10,000 g for 1 min
37. Discard the flow-through
38. Add 700 µl of DNA Wash Buffer 2 to the column
39. Centrifuge the Spin Column at 10,000 g for 1 min
40. Discard the flow-through
41. Add 200 µl of DNA Wash Buffer 2 to the column
42. Centrifuge the Spin Column at 10,000 g for 1 min
43. Transfer the column to a new, clean Elution Tube
44. Add 20 µl of DNase/RNase Free Water to the centre of the membrane of the MB Spin Column
45. Incubate the tube at room temperature for 5 min
46. Centrifuge the Spin Column at 10,000 g for 1 min
47. Add the flow-through again onto the centre of the membrane
48. Incubate the tube at room temperature for 5 min
49. Centrifuge the Spin Column at 10,000 g for 1 min
50. Store the extracted DNA at -20°C

Appendix 2 – R Code used for the analysis

```
#installing the required packages
if (!requireNamespace("BiocManager", quietly = TRUE)) install.packages("BiocManager")
BiocManager::install(c("phyloseq", "microbiome", "ComplexHeatmap"), update = FALSE)
install.packages("microViz", repos = c(davidbarnett = "https://david-barnett.r-
universe.dev", getOption("repos")))
install.packages("ggtext") #for rotated labels on ord_plot()
install.packages("ggraph") #for taxatree_plots()
install.packages("DT") #for tax_fix_interactive()
install.packages("corncob") #for beta binomial models in tax_model()
install.packages("tidyverse")
install.packages("funrar")
install.packages("devtools")
devtools::install_github("microsud/microbiomeutilities")
install.packages("patchwork")
install.packages("remotes")
remotes::install_github("schuyler-smith/phyloschuyler")
install.packages("gt")
install.packages("Hmisc")
remotes::install_github("adrientaudiere/MiscMetabar")
install.packages("BiocManager")
BiocManager::install("DESeq2")
install.packages("pheatmap)
install.packages("ggvenn")
install.packages("BSDA")
install.packages("usedist")
remotes::install_github("phytomosaic/ecole")

#loading the required packages
library(ggplot2)
library(phyloseq)
library(microViz)
library(plyr)
library(dplyr)
library(vegan)
library(tidyverse)
library(funrar)
library(microbiome)
library(knitr)
library(microbiomeutilities)
library("data.table")
library(patchwork)
library(scales)
library(gt)
library(grid)
library(reshape2)
library(Hmisc)
library("DESeq2")
library(pheatmap)
library(ggvenn)
library(BSDA)
library(usedist)
library(ecole)

#create a function to introduce line breaks in figure labels
wrap.it <- function(x, len) {sapply(x, function(y) paste(strwrap(y, len), collapse = "\n"),
USE.NAMES = FALSE)}
```

```
wrap.labels <- function(x, len) {if (is.list(x)) {lapply(x, wrap.it, len)} else {wrap.it(x,
len)}}
wrap.labels(colnames(tot_rel_abund_leaf_extraction), 10)

#read in the data file
final_data_16S <- readRDS("D:/Benutzer/Stefan/Downloads/JMF-2404-07_16S_phyloseq.rds")

#subset for my samples (there were other people's samples in this .rds file as well)
final_data_16S_SE <- subset_samples(final_data_16S,
str_detect(sample_data(final_data_16S)$User_sample_ID, "M-"))

#mark which samples were sequenced with PNAs
clamped <-
c(1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0)
sample_data(final_data_16S_SE)$clamped <- clamped

#mark which kit was used
kit_used <-
c("PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs
","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS",
"PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs","PS","PNAs",
"PS","QI","QI","QI","QI","QI","QI","QI","QI","QI","QI","QI","QI","QI","QI","QI","QI","QI","
QI","QI","QI","QI","QI","QI","HZ","HZ","HZ","HZ","HZ","HZ","HZ","HZ","HZ","HZ","HZ","HZ","H
Z","HZ","HZ","HZ","HZ","HZ","HZ","HZ","HZ","HZ","HZ")
sample_data(final_data_16S_SE)$kit_used <- kit_used

#repaires the dataset
final_data_16S_SE_fixed <- tax_fix(final_data_16S_SE)

#replaces all NA, unknown and empty names by the name in the next-higher phylogenetic level
final_data_16S_SE_newnames <- final_data_16S_SE_fixed %>% tax_rename(rank = "Species",
sort_by = prev, pad_digits = "max", sep = "-")

#prune data to include no taxa with 0 reads
final_data_16S_SE_newnames <- prune_taxa(taxa_sums(final_data_16S_SE_newnames) > 0,
final_data_16S_SE_newnames)

#subset data set for each method
final_data_16S_SE_clamped <- subset_samples(final_data_16S_SE_newnames,
str_detect(sample_data(final_data_16S_SE_newnames)$JMF_library_ID, "B"))
final_data_16S_SE_clamped_and_PS <- subset_samples(final_data_16S_SE_newnames,
str_detect(sample_data(final_data_16S_SE_newnames)$User_sample_ID, "-PS-"))
final_data_16S_SE_non_clamped <- subset_samples(final_data_16S_SE_newnames,
str_detect(sample_data(final_data_16S_SE_newnames)$JMF_library_ID, "C"))
final_data_16S_SE_PS <- subset_samples(final_data_16S_SE_non_clamped,
str_detect(sample_data(final_data_16S_SE_non_clamped)$User_sample_ID, "-PS-"))
final_data_16S_SE_QI <- subset_samples(final_data_16S_SE_newnames,
str_detect(sample_data(final_data_16S_SE_newnames)$User_sample_ID, "-QI-"))
final_data_16S_SE_HZ <- subset_samples(final_data_16S_SE_newnames,
str_detect(sample_data(final_data_16S_SE_newnames)$User_sample_ID, "-HZ-"))

#merge each set
final_data_16S_SE_merged <- merge_samples(final_data_16S_SE, "kit_used")
final_data_16S_merged_PS <- merge_samples(final_data_16S_SE_PS, "kit_used")
final_data_16S_SE_merged_fixed <- tax_fix(final_data_16S_SE_merged)
final_data_16S_SE_merged_newnames <- final_data_16S_SE_merged_fixed %>% tax_rename(rank =
"Species", sort_by = prev, pad_digits = "max", sep = "-")
```

```
comp_barplot(final_data_16S_SE_merged_newnames, "Family", n_taxa = 2) + coord_flip()
comp_barplot(final_data_16S_merged_PS, "Family", n_taxa = 2) + coord_flip()

final_data_16S_SE_merged_unclamped <- merge_samples(final_data_16S_SE_non_clamped,
"kit_used")
final_data_16S_SE_merged_fixed_unclamped <- tax_fix(final_data_16S_SE_merged_unclamped)
final_data_16S_SE_merged_newnames_unclamped <- final_data_16S_SE_merged_fixed_unclamped %>%
tax_rename(rank = "Species", sort_by = prev, pad_digits = "max", sep = "-")
comp_barplot(final_data_16S_SE_merged_newnames_unclamped, "Family", n_taxa = 2) +
coord_flip()

#subset to data set to include no host organelles
final_data_16S_SE_non_clamped_no_mito <- subset_taxa(final_data_16S_SE_non_clamped_without,
!Family=="Mitochondria")
final_data_16S_SE_non_clamped_no_mito_fixed <-
tax_fix(final_data_16S_SE_non_clamped_no_mito)
final_data_16S_SE_non_clamped_no_mito_newnames <-
final_data_16S_SE_non_clamped_no_mito_fixed %>% tax_rename(rank = "Species", sort_by =
prev, pad_digits = "max", sep = "-")
final_data_16S_SE_non_clamped_no_organelles <-
subset_taxa(final_data_16S_SE_non_clamped_no_mito_newnames, !Family=="Chloroplast Order")
final_data_16S_SE_non_clamped_no_organelles_fixed <-
tax_fix(final_data_16S_SE_non_clamped_no_organelles)
final_data_16S_SE_non_clamped_no_organelles_newnames <-
final_data_16S_SE_non_clamped_no_organelles_fixed %>% tax_rename(rank = "Species", sort_by
= prev, pad_digits = "max", sep = "-")

#merge the dataset without organelles
final_data_16S_SE_no_organelles_merged <-
merge_samples(final_data_16S_SE_non_clamped_no_organelles_newnames, "kit_used")
final_data_16S_SE_no_organelles_merged_fixed <-
tax_fix(final_data_16S_SE_no_organelles_merged)
final_data_16S_SE_no_organelles_merged_newnames <-
final_data_16S_SE_no_organelles_merged_fixed %>% tax_rename(rank = "Species", sort_by =
prev, pad_digits = "max", sep = "-")
comp_barplot(final_data_16S_SE_no_organelles_merged_newnames, "Family", n_taxa = 15) +
coord_flip()

#rarefy and merge the data
final_data_rare <- rarefy_even_depth(final_data_16S_SE_newnames, rngseed=111,
sample.size=8000, replace=F)
final_data_rare_merged_unfixed <- merge_samples(final_data_rare, "kit_used")
final_data_rare_merged_fixed <- tax_fix(final_data_rare_merged_unfixed)
final_data_rare_merged <- final_data_rare_merged_fixed %>% tax_rename(rank = "Species",
sort_by = prev, pad_digits = "max", sep = "-")
sample_names(final_data_rare_merged) <- c("PNAs", "HostZero", "Powersoil", "QIAamp")
final_data_rare_unclamped <- rarefy_even_depth(final_data_16S_SE_merged_newnames_unclamped,
rngseed=111, sample.size=8000, replace=F)
sample_names(final_data_rare_unclamped) <- c("HostZero", "Powersoil", "QIAamp")
final_data_rare_unclamped_unmerged <- rarefy_even_depth(final_data_16S_SE_non_clamped,
rngseed=111, sample.size=8000, replace=F)
```

Results

Host organelles dominate the sequenced seagrass leaf microbiome

```
#creating the barplot
comp_barplot(final_data_16S_merged_PS, "Family", n_taxa = 15, ylab="relative abundance",
label = "SAMPLE") + coord_flip() + theme(plot.title = element_text(hjust = 0.5, size = 15),
cex=1.2) + theme(axis.text.y = element_text(size = 15), legend.text = element_text(size =
15))
```

Host blocking with PNAs did not work well

```
#creating the barplots of host content with PNAs
par(mar=c(4,9.5,4,4), mfrow = c(2,1))
bp_overview_final_depletion_PNA_mito <- barplot(t(means_sd_table[1:2,1]), horiz = TRUE,
las=2, cex.names=1.2, cex.axis = 1.2, cex.lab = 1.2, xlim=c(0:1), xlab="relative
abundance", col=c("#C04F22"), names.arg=rownames(means_sd_table[2:4]), main="a)
mitochondrial content")
text(0.1, cex=1.2, bp_overview_final_depletion_PNA_mito, labels =
round(t(means_sd_table[1:2,1]), digits = 2))
segments(mean_mito[1:2] - sd_mito[1:2], bp_overview_final_depletion_PNA_mito,
mean_mito[1:2] + sd_mito[1:2], bp_overview_final_depletion_PNA_mito, lwd = 1.5)
arrows(mean_mito[1:2] - sd_mito[1:2], bp_overview_final_depletion_PNA_mito, mean_mito[1:2]
+ sd_mito[1:2], bp_overview_final_depletion_PNA_mito, lwd = 1.5, angle = 90, code = 3,
length = 0.05)
bp_overview_final_depletion_PNA_chloro <- barplot(t(means_sd_table[1:2,2]), horiz = TRUE,
las=2, cex.names=1.2, cex.axis = 1.2, cex.lab = 1.2, xlim=c(0:1), xlab="relative
abundance", col=c("#38CC3E"), names.arg=rownames(means_sd_table[2:4]), main="b) chloroplast
content")
text(0.1, cex=1.2, bp_overview_final_depletion_PNA_chloro, labels =
round(t(means_sd_table[1:2,2]), digits = 2))
segments(mean_chloro[1:2] - sd_chloro[1:2], bp_overview_final_depletion_PNA_chloro,
mean_chloro[1:2] + sd_chloro[1:2], bp_overview_final_depletion_PNA_chloro, lwd = 1.5)
arrows(mean_chloro[1:2] - sd_chloro[1:2], bp_overview_final_depletion_PNA_chloro,
mean_chloro[1:2] + sd_chloro[1:2], bp_overview_final_depletion_PNA_chloro, lwd = 1.5, angle
= 90, code = 3, length = 0.05)

#statistical testing
tsum.test(mean_mito_PS, sd_mito_PS, 19, mean_mito_PNA, sd_mito_PNA, 22, alternative =
"two.sided")
tsum.test(mean_chloro_PS, sd_chloro_PS, 19, mean_chloro_PNA, sd_chloro_PNA, 22, alternative
= "two.sided")
```

Host depletion with kits worked well

The QIAamp kit reduces host 'contamination' more efficiently than the HostZero kit

```
#creating the plot of the species composition
comp_barplot(final_data_rare_merged, "Family", n_taxa = 15, xlab="relative abundance",
label = "SAMPLE") + coord_flip() + ggtitle("a) species composition") + theme(plot.title =
element_text(hjust = 0.5, size = 15), cex=1.2) + theme(axis.text.y = element_text(size =
15), legend.text = element_text(size = 15))

comp_barplot(final_data_rare_unclamped, "Family", n_taxa = 15, xlab="relative abundance",
label = "SAMPLE") + coord_flip() + ggtitle("a) species composition") + theme(plot.title =
element_text(hjust = 0.5, size = 15), cex=1.2) + theme(axis.text.y = element_text(size =
15), legend.text = element_text(size = 15))
```

```
#make a table with the host content like before
final_data_rel_abund_rare <- prop.table(otu_table(final_data_rare_merged), margin = 1)
rel_abund_final_mito_rare <- final_data_rel_abund_rare[,grep("Mitoch",
colnames(final_data_rel_abund_rare))]
rel_abund_final_chloro_rare <- final_data_rel_abund_rare[,grep("Chloropl",
colnames(final_data_rel_abund_rare))]
tot_rel_mito_rare <- rowSums(rel_abund_final_mito_rare)
tot_rel_chloro_rare <- rowSums(rel_abund_final_chloro_rare)
rel_final_abund_table_host <- rbind (tot_rel_mito_rare, tot_rel_chloro_rare)
rownames(rel_final_abund_table_host) <- c("Mitochondria", "Chloroplasts")
mean_mito_HZ <- mean(rel_final_abund_table_host[1,1:16])
mean_chloro_HZ <- mean(rel_final_abund_table_host[2,1:16])
mean_mito_PNA <- mean(rel_final_abund_table_host[1,17:38])
mean_chloro_PNA <- mean(rel_final_abund_table_host[2,17:38])
mean_mito_PS <- mean(rel_final_abund_table_host[1,39:57])
mean_chloro_PS <- mean(rel_final_abund_table_host[2,39:57])
mean_mito_QI <- mean(rel_final_abund_table_host[1,58:66])
mean_chloro_QI <- mean(rel_final_abund_table_host[2,58:66])
sd_mito_HZ <- sd(rel_final_abund_table_host[1,1:16])
sd_chloro_HZ <- sd(rel_final_abund_table_host[2,1:16])
sd_mito_PNA <- sd(rel_final_abund_table_host[1,17:38])
sd_chloro_PNA <- sd(rel_final_abund_table_host[2,17:38])
sd_mito_PS <- sd(rel_final_abund_table_host[1,39:57])
sd_chloro_PS <- sd(rel_final_abund_table_host[2,39:57])
sd_mito_QI <- sd(rel_final_abund_table_host[1,58:66])
sd_chloro_QI <- sd(rel_final_abund_table_host[2,58:66])
mean_mito <- c(mean_mito_PNA, mean_mito_PS, mean_mito_HZ, mean_mito_QI)
mean_chloro <- c(mean_chloro_PNA, mean_chloro_PS, mean_chloro_HZ, mean_chloro_QI)
sd_mito <- c(sd_mito_PNA, sd_mito_PS, sd_mito_HZ, sd_mito_QI)
sd_chloro <- c(sd_chloro_PNA, sd_chloro_PS, sd_chloro_HZ, sd_chloro_QI)
means_sd_table <- cbind(mean_mito, mean_chloro, sd_mito, sd_chloro)
rownames(means_sd_table) <- c("PNAs", "PowerSoil", "HostZero", "QIAamp")
colnames(means_sd_table) <- c("Mitochondria", "Chloroplasts", "SD_Mitochondria",
"SD_Chloroplasts")

#creating the barplots of host content with the kits
par(mar=c(4,9.5,4,4), mfrow = c(2,1))
bp_overview_final_depletion_mito <- barplot(t(means_sd_table[2:4,1]), horiz = TRUE, las=2,
cex.names=1.2, cex.axis = 1.2, cex.lab = 1.2, xlim=c(0:1), legend=TRUE, xlab="relative
abundance", col=c("#C04F22"), names.arg=rownames(means_sd_table[2:4]), main="b)
mitochondrial content")
text(0.03, cex=1.2, bp_overview_final_depletion_mito, labels = round(means_sd_table[2:4,1],
digits = 2))
segments(mean_mito[2:4] - sd_mito[2:4], bp_overview_final_depletion_mito, mean_mito[2:4] +
sd_mito[2:4], bp_overview_final_depletion_mito, lwd = 1.5)
arrows(mean_mito[2:4] - sd_mito[2:4], bp_overview_final_depletion_mito, mean_mito[2:4] +
sd_mito[2:4], bp_overview_final_depletion_mito, lwd = 1.5, angle = 90, code = 3, length =
0.05)
bp_overview_final_depletion_chloro <- barplot(t(means_sd_table[2:4,2]), horiz = TRUE,
las=2, cex.names=1.2, cex.axis = 1.2, cex.lab = 1.2, xlim=c(0:1), legend=TRUE,
xlab="relative abundance", col=c("#38CC3E"), names.arg=rownames(means_sd_table[2:4]),
main="c) chloroplast content")
text(0.03, cex=1.2, bp_overview_final_depletion_chloro, labels =
round(means_sd_table[2:4,2], digits = 2))
segments(mean_chloro[2:4] - sd_chloro[2:4], bp_overview_final_depletion_chloro,
mean_chloro[2:4] + sd_chloro[2:4], bp_overview_final_depletion_chloro, lwd = 1.5)
arrows(mean_chloro[2:4] - sd_chloro[2:4], bp_overview_final_depletion_chloro,
mean_chloro[2:4] + sd_chloro[2:4], bp_overview_final_depletion_chloro, lwd = 1.5, angle =
90, code = 3, length = 0.05)
```

```
#statistical testing
tsum.test(mean_mito_PS, sd_mito_PS, 19, mean_mito_HZ, sd_mito_HZ, 16, alternative =
"two.sided")
tsum.test(mean_chloro_PS, sd_chloro_PS, 19, mean_chloro_HZ, sd_chloro_HZ, 16, alternative =
"two.sided")
tsum.test(mean_mito_PS, sd_mito_PS, 19, mean_mito_QI, sd_mito_QI, 9, alternative =
"two.sided")
tsum.test(mean_chloro_PS, sd_chloro_PS, 19, mean_chloro_QI, sd_chloro_QI, 9, alternative =
"two.sided")
```

Host depletion increases the number of prokaryotic groups detected

```
#creating the boxplots of alpha diversity indices
sample_data(final_data_rare_unclamped_unmerged)$Method <- c("Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"QIAamp", "QIAamp", "QIAamp", "QIAamp", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero")
plot_rich <- plot_richness(final_data_rare_unclamped_unmerged, x="Method",
measures=c("Observed", "Chao1", "Shannon"))
plot_rich + geom_boxplot() + geom_jitter(shape=16, position=position_jitter(0.2)) +
aes(x=Method, fill=Method)  + theme(legend.position="none")  +
scale_color_manual(values=c("HostZero"="#244D7F", "Powersoil"="#FD8E3F",
"QIAamp"="#6AB1D6")  + geom_point(aes(colour = Method), size = 4)) + theme(text =
element_text(size = 20), axis.text.x = element_text(angle = 90, hjust = 1)) +
theme(axis.title.x=element_blank()) + theme(axis.title.y=element_blank())

#statistical testing
estim_rich <- estimate_richness(final_data_rare_unclamped_unmerged, measures=c("Observed",
"Chao1", "Shannon"))
estim_rich <- t(estim_rich)
colnames(estim_rich) <- c("Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero")
estim_rich <- t(estim_rich)
estim_rich <- estim_rich[order(rownames(estim_rich)),]
mean_obs_HZ <- mean(estim_rich [1:16,1])
mean_obs_PS <- mean(estim_rich [39:57,1])
mean_obs_QI <- mean(estim_rich [58:66,1])
mean_chao_HZ <- mean(estim_rich [1:16,2])
mean_chao_PS <- mean(estim_rich [39:57,2])
mean_chao_QI <- mean(estim_rich [58:66,2])
mean_shan_HZ <- mean(estim_rich [1:16,4])
mean_shan_PS <- mean(estim_rich [39:57,4])
mean_shan_QI <- mean(estim_rich [58:66,4])
sd_obs_HZ <- sd(estim_rich [1:16,1])
sd_obs_PS <- sd(estim_rich [39:57,1])
sd_obs_QI <- sd(estim_rich [58:66,1])
sd_chao_HZ <- sd(estim_rich [1:16,2])
sd_chao_PS <- sd(estim_rich [39:57,2])
sd_chao_QI <- sd(estim_rich [58:66,2])
sd_shan_HZ <- sd(estim_rich [1:16,4])
```

```
sd_shan_PS <- sd(estim_rich [39:57,4])
sd_shan_QI <- sd(estim_rich [58:66,4])

#statistical testing
tsum.test(mean_obs_PS, sd_obs_PS, 19, mean_obs_HZ, sd_obs_HZ, 16, alternative =
"two.sided")
tsum.test(mean_obs_PS, sd_obs_PS, 19, mean_obs_QI, sd_obs_QI, 9, alternative = "two.sided")
tsum.test(mean_chao_PS, sd_chao_PS, 19, mean_chao_HZ, sd_chao_HZ, 16, alternative =
"two.sided")
tsum.test(mean_chao_PS, sd_chao_PS, 19, mean_chao_QI, sd_chao_QI, 9, alternative =
"two.sided")
tsum.test(mean_shan_PS, sd_shan_PS, 19, mean_shan_HZ, sd_shan_HZ, 16, alternative =
"two.sided")
tsum.test(mean_shan_PS, sd_shan_PS, 19, mean_shan_QI, sd_shan_QI, 9, alternative =
"two.sided")
```

Host depletion introduces biases into the sequencing results

There were large differences in which taxa were detected by which kit

```
#creating the venn diagram
final_data_organel_1 <- subset_taxa(final_data_rare, !Family== "Mitochondria")
final_data_organel <- subset_taxa(final_data_organel_1, !Family== "Chloroplast Order")
final_data_rare_df <- psmelt(final_data_16S_SE_non_clamped_no_organelles_newnames)
final_data_rare_number_taxa <- final_data_rare_df %>% filter(Abundance > 0) %>%
group_by(kit_used, Genus) %>% summarise(n = length(Abundance))
venn_list_HZ <-
unique(final_data_rare_number_taxa$Genus[final_data_rare_number_taxa$kit_used == "HZ"])
venn_list_PS <-
unique(final_data_rare_number_taxa$Genus[final_data_rare_number_taxa$kit_used == "PS"])
venn_list_QI <-
unique(final_data_rare_number_taxa$Genus[final_data_rare_number_taxa$kit_used == "QI"])
venn_list_all <- list("HostZero" = venn_list_HZ, "PowerSoil" = venn_list_PS, "QIAamp" =
venn_list_QI)
ggvenn(venn_list_all, c("HostZero", "PowerSoil", "QIAamp"), fill_color = c("#244D7F",
"#FD8E3F", "#6AB1D6"))
```

Community differences induced by host depletion are only partially due to the removal of host organelles

```
#creating the NMDS including the host organelles
final_data_rare_ord_unclamped <- ordinate(final_data_rare_unclamped_unmerged, "NMDS",
"bray")
nmds_final_data_nonmerged_rare_unclamped <-
plot_ordination(final_data_rare_unclamped_unmerged, final_data_rare_ord_unclamped,
type="samples", color="Method") + geom_point(size=5) +
scale_color_manual(values=c("HostZero"="#244D7F", "Powersoil"="#FD8E3F",
"QIAamp"="#6AB1D6")) + stat_ellipse(geom = "polygon", alpha = 0.05, lwd=1.2) +
theme_bw(base_size = 18) + xlab("NMDS1") + ylab("NMDS2") + ggtitle(label="a) data
containing the host organelles")
print(nmds_final_data_nonmerged_rare_unclamped)

#statistical analysis
dist_organel <- phyloseq::distance(final_data_rare_unclamped_unmerged, method = "bray")
dist_organel <- dist_setNames(dist_organel, c("Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
```

```
"Powersoil", "Powersoil", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"QIAamp", "QIAamp", "QIAamp", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero"))
metadata_organel <- as(sample_data(final_data_rare_unclamped_unmerged), "data.frame")
permanova_pairwise(dist_organel, metadata_organel$kit_used, permutations=9999)

#creating the NMDS not including the host organelles
final_data_prokary <- final_data_16S_SE_non_clamped_no_organelles_newnames
final_data_ord_prokary_NMDS <- ordinate(final_data_prokary, "NMDS", "bray")
nmds_final_data_prokary <- plot_ordination(final_data_prokary, final_data_ord_prokary_NMDS,
type="samples", color="kit_used") + geom_point(size=5) +
scale_color_manual(values=c("HZ"="#244D7F", "PS"="#FD8E3F", "QI"="#6AB1D6")) +
stat_ellipse(geom = "polygon", alpha = 0.05, lwd=1.2) + theme_bw(base_size = 18) +
xlab("NMDS1") + ylab("NMDS2") + ggtitle(label="b) data not containing the host organelles")
print(nmds_final_data_prokary)

#statistical analysis
dist_prokary <- phyloseq::distance(final_data_prokary, method = "bray")
dist_prokary <- dist_setNames(dist_prokary, c("Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"QIAamp", "QIAamp", "QIAamp", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero"))
metadata_prokary <- as(sample_data(final_data_prokary), "data.frame")
permanova_pairwise(dist_prokary, metadata_prokary$kit_used, permutations=9999)
```

Host depletion kits deplete archaeal abundance along with host 'contamination'

```
#archaeal content
final_data_16S_domain <- tax_agg(final_data_rare, rank = "Domain")
final_data_rel_abund_domain <- prop.table(otu_table(final_data_16S_domain), margin = 1)
colnames(final_data_rel_abund_domain) <- c("PNA", "Powersoil", "PNA", "Powersoil", "PNA",
"Powersoil", "PNA", "Powersoil", "PNA", "Powersoil", "PNA", "Powersoil", "PNA",
"Powersoil", "PNA", "Powersoil", "PNA", "Powersoil", "Powersoil", "PNA", "Powersoil",
"PNA", "Powersoil", "PNA", "Powersoil", "PNA", "Powersoil", "PNA", "PNA", "Powersoil",
"PNA", "PNA", "Powersoil", "PNA", "Powersoil", "PNA", "PNA", "Powersoil", "PNA", "PNA",
"Powersoil", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"QIAamp", "QIAamp", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero")
final_data_rel_abund_domain <-
final_data_rel_abund_domain[,order(colnames(final_data_rel_abund_domain))]
mean_archaea_HZ <- mean(final_data_rel_abund_domain[2,1:16])
mean_archaea_PNA <- mean(final_data_rel_abund_domain[2,17:38])
mean_archaea_PS <- mean(final_data_rel_abund_domain[2,39:57])
mean_archaea_QI <- mean(final_data_rel_abund_domain[2,58:66])
sd_archaea_HZ <- sd(final_data_rel_abund_domain[2,1:16])
sd_archaea_PNA <- sd(final_data_rel_abund_domain[2,17:38])
sd_archaea_PS <- sd(final_data_rel_abund_domain[2,39:57])
sd_archaea_QI <- sd(final_data_rel_abund_domain[2,58:66])
mean_archaea <- cbind(mean_archaea_HZ, mean_archaea_PNA, mean_archaea_PS, mean_archaea_QI)
sd_archaea <- cbind(sd_archaea_HZ, sd_archaea_PNA, sd_archaea_PS, sd_archaea_QI)
table_overview_archaea <- rbind(mean_archaea, sd_archaea)
colnames(table_overview_archaea) <- c("HostZero", "PNAs", "PowerSoil", "QIAamp")
rownames(table_overview_archaea) <- c("mean", "SD")
```

```
table_overview_archaea = subset(table_overview_archaea, select = -c(PNAs) )

#making the barplot
par(mar=c(4,9.5,4,4), mfrow = c(1,1))
bp_overview_final_archaea <- barplot(table_overview_archaea[1,], horiz = TRUE, las=2,
xlab="relative abundance", col=c("#41B7C4"), names.arg=colnames(table_overview_archaea),
xlim=c(0,0.05), cex.names=1.2, cex.axis = 1.2, cex.lab = 1.2)
text(0.0015, bp_overview_final_archaea, labels = round(table_overview_archaea[1,], digits =
3), cex=1.2)
arrows(table_overview_archaea[1,] - table_overview_archaea[2,], bp_overview_final_archaea,
table_overview_archaea[1,] + table_overview_archaea[2,], bp_overview_final_archaea, lwd =
1.5, angle = 90, code = 3, length = 0.05)
segments(table_overview_archaea[1,] - table_overview_archaea[2,],
bp_overview_final_archaea, table_overview_archaea[1,] + table_overview_archaea[2,],
bp_overview_final_archaea, lwd = 1.5)

#statistical testing
tsum.test(mean_archaea_PS, sd_archaea_PS, 19, mean_archaea_PNA , sd_archaea_PNA, 22,
alternative = "two.sided")
tsum.test(mean_archaea_PS, sd_archaea_PS, 19, mean_archaea_HZ, sd_archaea_HZ, 16,
alternative = "two.sided")
tsum.test(mean_archaea_PS, sd_archaea_PS, 19, mean_archaea_QI, sd_archaea_QI, 9,
alternative = "two.sided")
```

Host depletion enabled deeper analysis of the seagrass microbiome

```
#creating the plot to compare PS vs QIA
final_data_comparision <- subset_samples(final_data_16S_SE_non_clamped,
!str_detect(sample_data(final_data_16S_SE_non_clamped)$kit_used, "HZ"))
dds_PS_QIA <- phyloseq_to_deseq2(final_data_comparision, ~ kit_used)
dds_PS_QIA <- DESeq(dds_PS_QIA, test="Wald", fitType="parametric")
res_dds_PS_QIA <- results(dds_PS_QIA, cooksCutoff = FALSE)
alpha_dds_PS_QIA <- 0.01
sigtab <- res_dds_PS_QIA[which(res_dds_PS_QIA$padj < alpha_dds_PS_QIA), ]
sigtab  <- cbind(as(sigtab , "data.frame"),
as(tax_table(final_data_comparision)[rownames(sigtab), ], "matrix"))
head(sigtab)
theme_set(theme_bw())
scale_fill_discrete <- function(palname = "Set1", ...) {
    scale_fill_brewer(palette = palname, ...)
}
x <- tapply(sigtab$log2FoldChange, sigtab$Genus, function(x) max(x))
x <- sort(x, TRUE)
sigtab$Genus = factor(as.character(sigtab$Genus), levels=names(x))
ggplot(sigtab, aes(x=Genus, y=log2FoldChange, color=Phylum)) + geom_point(size=1.5) +
theme(axis.text.x = element_text(angle = -90, hjust = 0, vjust = 0.5)) + theme(axis.text.y
= element_text(size = 7), axis.title.y = element_blank(), legend.text =
element_text(size=7), legend.title = element_text(size = 9), axis.title.x =
element_text(size=9), legend.position = "none")
final_data_glom <- tax_glom(final_data_16S_SE_rare, taxrank = 'Genus')
final_data_glom_rel <- abundances(final_data_glom, "compositional")
```

Quality control of sequencing

```
#creating the barplot of DNA yield
final_data_DNA_all <- read.csv("D:/Benutzer/Stefan/Dokumente/Uni/MSc/V. Masterarbeit/Arbeit
selber/2024-05-03 Sample Data JMF.csv", header = TRUE, sep=";")
final_data_DNA <- data.frame(ID=c("M-PS-L", "M-QI-L", "M-HZ-L"),
concentration=c(mean(final_data_DNA_all[1:23,5]), mean(final_data_DNA_all[24:46,5]),
mean(final_data_DNA_all[47:69,5])),stringsAsFactors = FALSE)
means_DNA <- final_data_DNA[,2]
sd_DNA <- c(sd(final_data_DNA_all[1:23,2]), sd(final_data_DNA_all[24:46,2]),
sd(final_data_DNA_all[47:69,2]))
par(mar=c(4,9.5,4,4), mfrow = c(1,1))
bp_overview_final_DNA <- barplot(final_data_DNA[,2], horiz=TRUE, cex.names=1.2,
col=c("#c6b3d8"), xlab="DNA yield [ng/mg biomass]", names.arg=c("PowerSoil", "QIAamp",
"HostZero"), las=2, xlim=c(0,28))
segments(means_DNA - sd_DNA, bp_overview_final_DNA, means_DNA + sd_DNA ,
bp_overview_final_DNA, lwd = 1.5)
arrows(means_DNA - sd_DNA, bp_overview_final_DNA, means_DNA + sd_DNA,
bp_overview_final_DNA, lwd = 1.5, angle = 90, code = 3, length = 0.05)
text(5, bp_overview_final_DNA, labels = round(means_DNA, digits = 2))

#creating the barplot of average read count
readcount_final <- sample_data(final_data_16S_SE_newnames)$reads_sample
names(readcount_final) <- sample_data(final_data_rare)$kit_used
readcount_final <- readcount_final[order(names(readcount_final))]
mean_read_HZ <- mean(readcount_final[1:16])
mean_read_PNA <- mean(readcount_final[17:38])
mean_read_PS <- mean(readcount_final[39:57])
mean_read_QI <- mean(readcount_final[58:66])
sd_read_HZ <- sd(readcount_final[1:16])
sd_read_PNA <- sd(readcount_final[17:38])
sd_read_PS <- sd(readcount_final[39:57])
sd_read_QI <- sd(readcount_final[58:66])
mean_read_final <- cbind(mean_read_HZ, mean_read_PNA, mean_read_PS, mean_read_QI)
sd_read_final <- cbind(sd_read_HZ, sd_read_PNA, sd_read_PS, sd_read_QI)
read_final <- rbind(mean_read_final, sd_read_final)
colnames(read_final) <- c("HostZero","PNAs","PowerSoil","QIAamp")
rownames(read_final) <- c("mean","SD")
par(mar=c(4,9.5,4,4), mfrow = c(1,1))
bp_readcount <- barplot(read_final[1,], horiz=TRUE, col=c("#90ee90"), xlab="read count",
names.arg=wrap.labels(colnames(read_final), 12), las=2, xlim=c(0,13000))
segments(read_final[1,] - read_final[2,], bp_readcount, read_final[1,] + read_final[2,],
bp_readcount, lwd = 1.5)
arrows(read_final[1,] - read_final[2,], bp_readcount, read_final[1,] + read_final[2,],
bp_readcount, lwd = 1.5, angle = 90, code = 3, length = 0.05)
text(1500, bp_readcount, labels = round(read_final[1,], digits = 0))

#statistical testing
tsum.test(mean_read_PS, sd_read_PS, 19, mean_read_HZ, sd_read_HZ, 16, alternative =
"two.sided")
tsum.test(mean_read_PS, sd_read_PS, 19, mean_read_QI, sd_read_QI, 9, alternative =
"two.sided")
tsum.test(mean_read_HZ, sd_read_HZ, 16, mean_read_QI, sd_read_QI, 9, alternative =
"two.sided")
tsum.test(mean_read_PNA, sd_read_PNA, 22, mean_read_HZ, sd_read_HZ, 16, alternative =
"two.sided")
tsum.test(mean_read_PNA, sd_read_PNA, 22, mean_read_QI, sd_read_QI, 9, alternative =
"two.sided")
```

```
tsum.test(mean_read_PNA, sd_read_PNA, 22, mean_read_PS, sd_read_PS, 19, alternative =
"two.sided")

#creating the histogram of read length
par(mar=c(4,7,4,4), mfrow = c(1,2))
hist(sample_sums(final_data_16S_SE_newnames), main="a) histogram: read counts", xlab="Total
Reads", border="darkblue", col="lightgreen", las=1, breaks=12)
final_data_reads <- readcount(final_data_16S_SE_newnames)
names(final_data_reads) <- sample_data(final_data_16S_SE_newnames)$User_sample_ID
reads_final_data <- sort(final_data_reads)
sample_data(final_data_16S_SE_newnames)$reads_sample <- final_data_reads
bp_reads_final_data <- barplot(reads_final_data, xlab = "number of reads", las=1,
cex.names=0.5, horiz=TRUE, main = "b) read counts", border="darkblue", col="lightgreen")
abline(v=8000)

#creating the rarefaction curves
par(mar=c(4,4,4,4), mfrow = c(1,1))
otu_rare_final_data <- as.data.frame(t(otu_table(final_data_16S_SE_newnames)))
otu_rarecurve_final <- rarecurve(otu_rare_final_data, step = 100, label = FALSE, col =
"darkgreen", main = "c) rarefaction curve")
abline(v=8000)
```

Supplementary Results

α-diversity behaved the same across all different indices

```
#creating the boxplots of all alpha diversity indices
plot_rich_all <- plot_richness(final_data_rare_unclamped_unmerged, x="Method",
measures=NULL)
plot_rich_all + geom_boxplot() + geom_jitter(shape=16, position=position_jitter(0.2)) +
aes(x=Method, fill=Method)  + theme(legend.position="none")  +
scale_color_manual(values=c("HostZero"="#244D7F", "Powersoil"="#FD8E3F",
"QIAamp"="#6AB1D6")  + geom_point(aes(colour = Method), size = 4)) + theme(text =
element_text(size = 20), axis.text.x = element_text(angle = 90, hjust = 1)) +
theme(axis.title.x=element_blank()) + theme(axis.title.y=element_blank())

#statistical testing
estim_rich_all <- estimate_richness(final_data_rare_unclamped_unmerged, measures=NULL)
estim_rich_all <- t(estim_rich_all)
colnames(estim_rich_all) <- c("Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"QIAamp", "QIAamp", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero")
estim_rich_all <- t(estim_rich_all)
estim_rich_all <- estim_rich_all[order(rownames(estim_rich_all)),]
mean_ace_HZ <- mean(estim_rich_all[1:16,4])
mean_ace_PS <- mean(estim_rich_all[17:35,4])
mean_ace_QI <- mean(estim_rich_all[36:44,4])
sd_ace_HZ <- sd(estim_rich_all[1:16,4])
sd_ace_PS <- sd(estim_rich_all[17:35,4])
sd_ace_QI <- sd(estim_rich_all[36:44,4])
mean_simp_HZ <- mean(estim_rich_all[1:16,7])
mean_simp_PS <- mean(estim_rich_all[17:35,7])
mean_simp_QI <- mean(estim_rich_all[36:44,7])
sd_simp_HZ <- sd(estim_rich_all[1:16,7])
```

```
sd_simp_PS <- sd(estim_rich_all[17:35,7])
sd_simp_QI <- sd(estim_rich_all[36:44,7])
mean_insimp_HZ <- mean(estim_rich_all[1:16,8])
mean_insimp_PS <- mean(estim_rich_all[17:35,8])
mean_insimp_QI <- mean(estim_rich_all[36:44,8])
sd_insimp_HZ <- sd(estim_rich_all[1:16,8])
sd_insimp_PS <- sd(estim_rich_all[17:35,8])
sd_insimp_QI <- sd(estim_rich_all[36:44,8])
mean_fish_HZ <- mean(estim_rich_all[1:16,9])
mean_fish_PS <- mean(estim_rich_all[17:35,9])
mean_fish_QI <- mean(estim_rich_all[36:44,9])
sd_fish_HZ <- sd(estim_rich_all[1:16,9])
sd_fish_PS <- sd(estim_rich_all[17:35,9])
sd_fish_QI <- sd(estim_rich_all[36:44,9])
tsum.test(mean_ace_PS, sd_ace_PS, 19, mean_ace_HZ, sd_ace_HZ, 16, alternative =
"two.sided")
tsum.test(mean_ace_PS, sd_ace_PS, 19, mean_ace_QI, sd_ace_QI, 9, alternative = "two.sided")
tsum.test(mean_simp_PS, sd_simp_PS, 19, mean_simp_HZ, sd_simp_HZ, 16, alternative =
"two.sided")
tsum.test(mean_simp_PS, sd_simp_PS, 19, mean_simp_QI, sd_simp_QI, 9, alternative =
"two.sided")
tsum.test(mean_insimp_PS, sd_insimp_PS, 19, mean_insimp_HZ, sd_insimp_HZ, 16, alternative =
"two.sided")
tsum.test(mean_insimp_PS, sd_insimp_PS, 19, mean_insimp_QI, sd_insimp_QI, 9, alternative =
"two.sided")
tsum.test(mean_fish_PS, sd_fish_PS, 19, mean_fish_HZ, sd_fish_HZ, 16, alternative =
"two.sided")
tsum.test(mean_fish_PS, sd_fish_PS, 19, mean_fish_QI, sd_fish_QI, 9, alternative =
"two.sided")
```

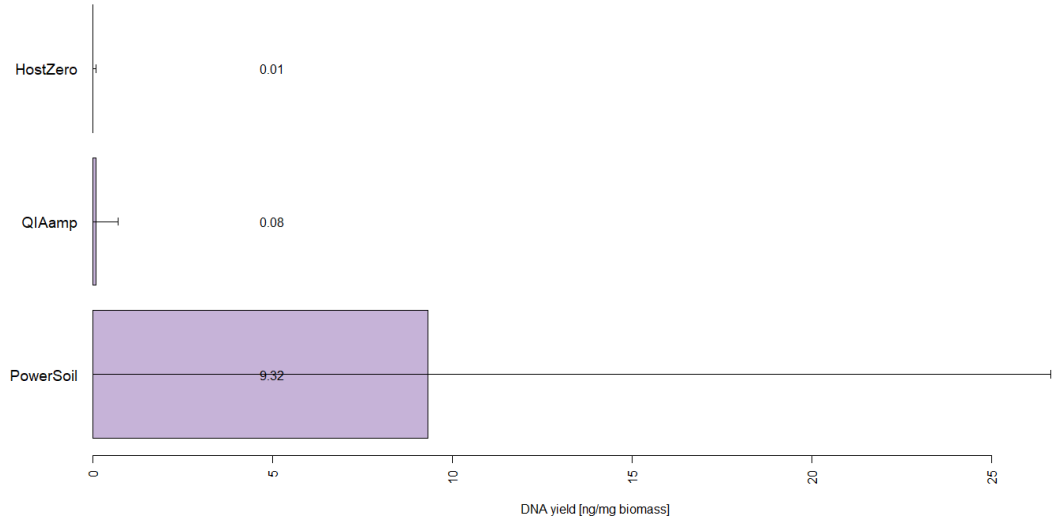
All β-diversity analyses showed similar results

```
#creating the PCoA including the host organelles
final_data_ord_pcoa <- ordinate(final_data_rare_unclamped_unmerged, "PCoA", "bray")
pcoa_final_data_ord <- plot_ordination(final_data_rare_unclamped_unmerged,
final_data_ord_pcoa, type="samples", color="kit_used") + geom_point(size=5) +
scale_color_manual(values=c("HZ"="#244D7F", "PS"="#FD8E3F", "QI"="#6AB1D6")) + theme(text =
element_text(size = 15)) + stat_ellipse(geom = "polygon", alpha = 0.05, lwd=1.2) +
theme_bw(base_size = 18) + ggtitle(label="a) data containing the host organelles")
print(pcoa_final_data_ord)

#statistical analysis
dist_organel_pcoa <- phyloseq::distance(final_data_rare_unclamped_unmerged, method =
"bray")
dist_organel_pcoa <- dist_setNames(dist_organel, c("Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"QIAamp", "QIAamp", "QIAamp", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero"))
metadata_organel_pcoa <- as(sample_data(final_data_rare_unclamped_unmerged), "data.frame")
permanova_pairwise(dist_organel_pcoa, metadata_organel_pcoa$kit_used, permutations=9999)

#creating the PCoA not including the host organelles
final_data_prokary_pcoa <- ordinate(final_data_prokary, "PCoA", "bray")
pcoa_final_prokary <- plot_ordination(final_data_prokary, final_data_prokary_pcoa,
type="samples", color="kit_used") + geom_point(size=5) +
```

```
scale_color_manual(values=c("HZ"="#244D7F", "PS"="#FD8E3F", "QI"="#6AB1D6")) + theme(text =
element_text(size = 15)) + stat_ellipse(geom = "polygon", alpha = 0.05, lwd=1.2) +
theme_bw(base_size = 18) + ggtitle(label="b) data not containing the host organelles")
print(pcoa_final_prokary)

#statistical analysis
dist_prokary_pcoa <- phyloseq::distance(final_data_prokary, method = "bray")
dist_prokary_pcoa <- dist_setNames(dist_prokary, c("Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"QIAamp", "QIAamp", "QIAamp", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero"))
metadata_prokary_pcoa <- as(sample_data(final_data_prokary), "data.frame")
permanova_pairwise(dist_prokary_pcoa, metadata_prokary_pcoa$kit_used, permutations=9999)

#creating the CAP including the host organelles
final_data_rare_unclamped_unmerged_bray <-
phyloseq::distance(final_data_rare_unclamped_unmerged, method = "bray")
cap_ord_rare <- ordinate(final_data_rare_unclamped_unmerged, "CAP",
final_data_rare_unclamped_unmerged_bray, formula = ~ kit_used)
cap_plot_rare <- plot_ordination(final_data_rare, cap_ord_rare, color = "kit_used", axes =
c(1,2)) + geom_point(aes(colour = kit_used), size = 4) + geom_point(size = 1.5) +
scale_color_manual(values=c("HZ"="#244D7F", "PS"="#FD8E3F", "QI"="#6AB1D6")) +
theme_bw(base_size = 18)
arrowmat <- vegan::scores(cap_ord_rare, display = "cn")
rownames(arrowmat) <- c("HostZero", "PowerSoil", "QIAamp")
arrowdf <- data.frame(labels = rownames(arrowmat), arrowmat)
arrow_map <- aes(xend = CAP1, yend = CAP2, x = 0, y = 0, shape = NULL, color = kit_used,
label = rownames(arrowmat))
label_map <- aes(x = 1.3 * CAP1, y = 1.3 * CAP2, shape = NULL, color = NULL, label =
rownames(arrowmat))
arrowhead = arrow(length = unit(0.02, "npc"))
cap_plot_rare + geom_segment(mapping = arrow_map, size = .9, data = arrowdf, color =
"black", arrow = arrowhead) + geom_text(mapping = label_map, size = 4.8,  data = arrowdf,
show.legend = TRUE) + stat_ellipse(geom = "polygon", alpha = 0.05, lwd=1.2) +
ggtitle(label="a) data containing the host organelles")

#statistical analysis
dist_final_data_rare_unclamped_unmerged_cap <-
phyloseq::distance(final_data_rare_unclamped_unmerged, method = "bray")
dist_final_data_rare_unclamped_unmerged_cap <-
dist_setNames(dist_final_data_rare_unclamped_unmerged_cap, c("Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"QIAamp", "QIAamp", "QIAamp", "QIAamp", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero"))
metadata_final_data_rare_unclamped_unmerged_cap <-
as(sample_data(final_data_rare_unclamped_unmerged), "data.frame")
permanova_pairwise(dist_final_data_rare_unclamped_unmerged_cap,
metadata_final_data_rare_unclamped_unmerged_cap$kit_used, permutations=9999)

#creating the CAP not including the host organelles
final_data_prokary_bray <- phyloseq::distance(final_data_prokary, method = "bray")
cap_ord_rare_prokary <- ordinate(final_data_prokary, "CAP", final_data_prokary_bray,
formula = ~ kit_used)
```

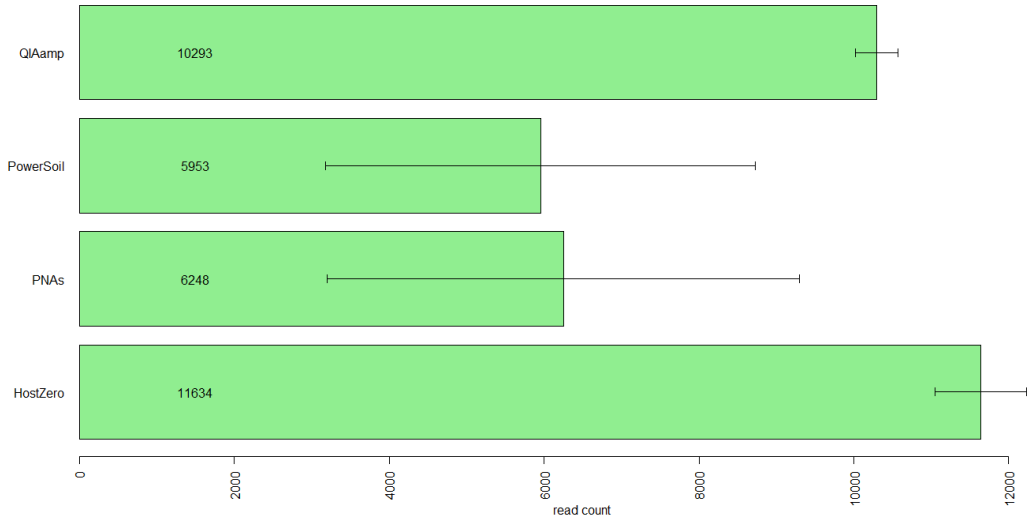
```
cap_plot_prokary <- plot_ordination(final_data_prokary, cap_ord_rare_prokary, color =
"kit_used", axes = c(1,2)) + geom_point(aes(colour = kit_used), size = 4) + geom_point(size
= 1.5) + scale_color_manual(values=c("HZ"="#244D7F", "PS"="#FD8E3F", "QI"="#6AB1D6")) +
theme_bw(base_size = 18)
arrowmat <- vegan::scores(cap_ord_rare_prokary, display = "cn")
rownames(arrowmat) <- c("HostZero", "PowerSoil", "QIAamp")
arrowdf <- data.frame(labels = rownames(arrowmat), arrowmat)
arrow_map <- aes(xend = CAP1, yend = CAP2, x = 0, y = 0, shape = NULL, color = kit_used,
label = rownames(arrowmat))
label_map <- aes(x = 1.3 * CAP1, y = 1.3 * CAP2, shape = NULL, color = NULL, label =
rownames(arrowmat))
arrowhead = arrow(length = unit(0.02, "npc"))
cap_plot_prokary + geom_segment(mapping = arrow_map, size = .9, data = arrowdf, color =
"black", arrow = arrowhead) + geom_text(mapping = label_map, size = 4.8,  data = arrowdf,
show.legend = TRUE) + stat_ellipse(geom = "polygon", alpha = 0.05, lwd=1.2) +
ggtitle(label="b) data not containing the host organelles")

#statistical analysis
dist_prokary_cap <- phyloseq::distance(final_data_prokary, method = "bray")
dist_prokary_cap <- dist_setNames(dist_prokary_cap, c("Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil", "Powersoil",
"Powersoil", "Powersoil", "Powersoil", "QIAamp", "QIAamp", "QIAamp", "QIAamp", "QIAamp",
"QIAamp", "QIAamp", "QIAamp", "QIAamp", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero",
"HostZero", "HostZero", "HostZero", "HostZero", "HostZero", "HostZero"))
metadata_prokary_cap <- as(sample_data(final_data_prokary), "data.frame")
permanova_pairwise(dist_prokary_cap, metadata_prokary_cap$kit_used, permutations=9999)
```

Appendix 3 – Quality control of sequencing



Supplementary Figure 1. ***DNA yield of samples treated with different host depletion kits.*** *The x-axis depicts the DNA yield normalised to milligram of biomass centration in the different samples treated with different host depletion methods listed on the y-axis. The y-axis lists the kit that was used for host depletion. The numbers in each bar signify the exact value for that sample. The error bars denote standard deviation.*

I performed both DNA extraction and host depletion according to the previously optimised protocols. I measured DNA yield. Then, for further analysis, I sequenced the 16S rRNA of all samples, using mitochondrial and chloroplast reads as proxy for host content. After this, I evaluated the read count distribution to see how well the sequencing worked. Finally, I performed rarefaction to make all the different samples comparable to each other.

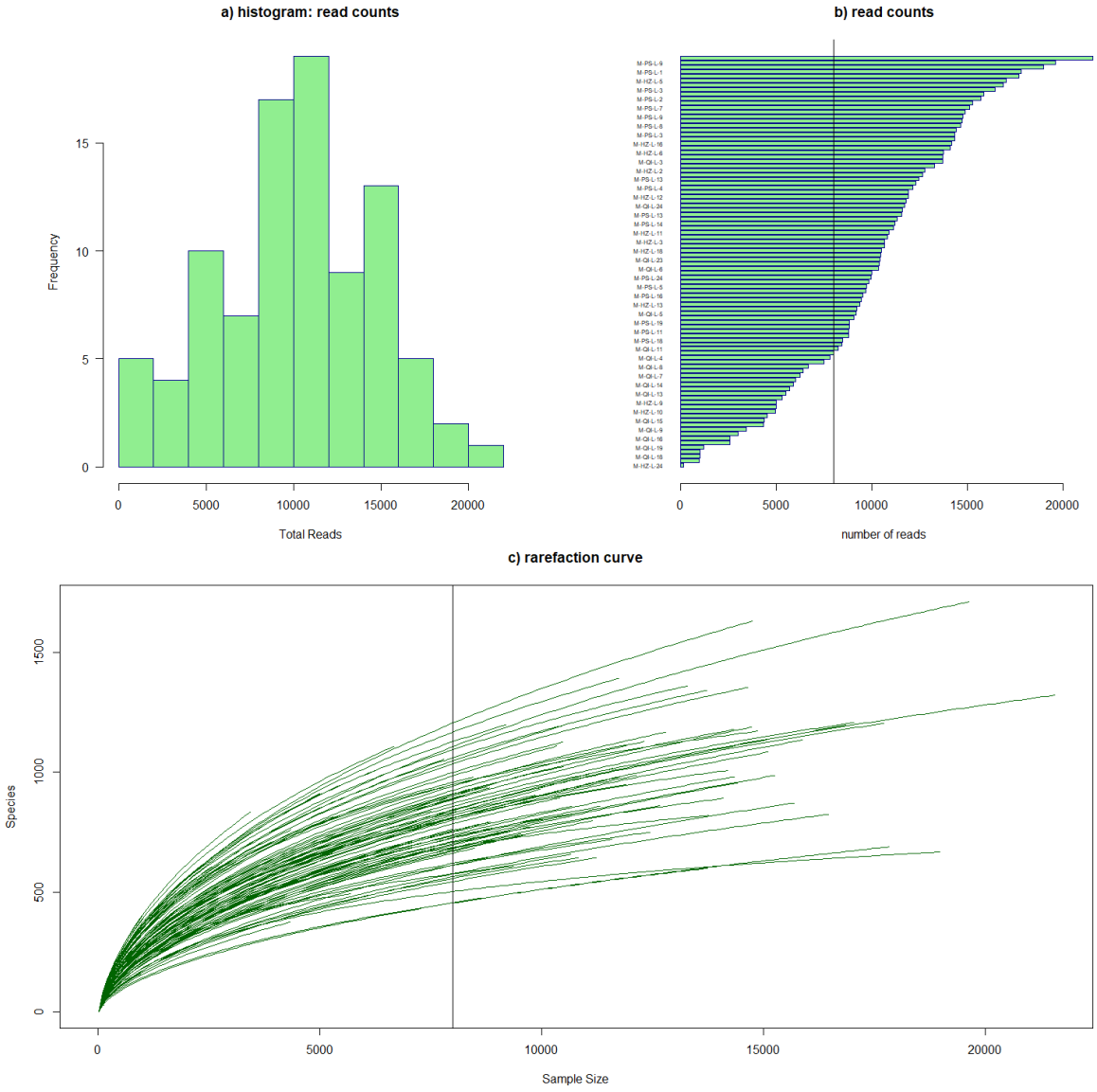


Supplementary Figure 2. ***Read length of samples treated with different host depletion kits.*** *The x-axis depicts the read count in the different samples treated with different host depletion methods listed on the y-axis. The y-axis lists the kit that was used for host depletion. The numbers in each bar signify the exact value for that sample. The error bars denote standard deviation. Neither HostZero nor QIAamp depletion significantly decreased the read count, both increased it significantly (HostZero* $p=3.205x10^{-8}$*, QIAamp* $p=2.017x10^{-6}$*).*

DNA concentration after extraction was extremely different between the three methods used. Extraction with PowerSoil yielded by far the highest concentration, with 9,32 ng DNA per mg of biomass. The two host depletion kits yielded DNA with concentrations lower by orders of magnitude.

QIAamp yielded only 0.08 ng DNA per mg biomass, and HostZero even less at 0.01 ng DNA per mg biomass (Supplementary Figure 1.). In the HostZero host depletion kit, there were eleven samples whose concentration was given as too low for our measurement, meaning below 0.1 ng/µl. Such samples were not present for any of the other kits.

Even though the DNA yield was low, it was still possible to sequence those samples. In fact, the number of reads generated by sequencing was not at all correlated to the DNA yield. HostZero, the kit with the lowest DNA yield (Supplementary Figure 1.) had the highest average read count at more than 11,000 reads (Supplementary Figure 2.), statistically significantly higher than the samples extracted with the PowerSoil kit at about 6,000 reads. QIAamp also had a high read count, more than 10,000 reads, that was statistically significantly higher than the PowerSoil samples (Supplementary Figure 2.).



Supplementary Figure 3. ***Quality control of sequencing results of the final comparison.*** *The sequencing quality of the DNA samples from* Zostera marina *after DNA extraction and/or treatment with different host depletion kits. a) shows the histogram for the number of reads sequenced for each sample, the x-axis represents the total number of reads, the y-axis depicts the frequency of that number of reads. b) shows the exact read count for every sample, the x-axis depicts the number of reads. The vertical line marks the cut-off for rarefaction, 8000 reads. c) shows the rarefaction curve for every sample, the x-axis shows the number of reads samples, the y-axis shows the number of species detected in those reads. The vertical line marks the cut-off for rarefaction, 8000 reads.*

After sequencing, the read count was evaluated. The number of reads was spread out between 100 and 22,000 reads (Supplementary Figure 3.). The lowest number of reads was 135, observed in one of the samples treated with the HostZero kit which had had a DNA concentration that was below the detection threshold. The highest read count was 21,563, in one of the samples extracted using the PowerSoil kit and then sequenced with PCR inhibitors. It had a DNA concentration of 28.0 ng/µl, which was not among the highest concentrations observed, it was in the lower half of concentrations observed for the samples extracted with PowerSoil (Supplementary Figure 3.).

The rarefaction curves for all samples looked as should have been expected. While none of the samples were saturated, most of them were already starting to dip down towards the end, except for those samples that had very low read counts, meaning that a large part of the community present was already captured by this sequencing depth (Supplementary Figure 3.). Because of this, I decided not to discard any samples based on sequencing quality, but to continue the analysis with all of them that had a satisfying number of sequencing reads. To enable better comparison between the different samples, I rarefied all of them to a read count of 8000. This reduced the number of samples from 92 to 66, which is a loss of 28.3% of all those samples. Of the remaining samples, 19 samples had been extracted with the PowerSoil kit and sequenced without PNAs, 22 samples had been extracted with the PowerSoil kit and sequenced with PNAs, 9 samples had been extracted with the QIAamp kit, and 16 samples with the HostZero kit. This meant a loss of 17.4% of all samples extracted with the PowerSoil kit and sequenced without PNAs, 4.3% of all samples extracted with the PowerSoil kit and sequenced with PNAs, a loss of 60.9% of all samples extracted with the QIAamp kit, and a loss of 30.4% of all samples extracted with the HostZero kit. But after all, the number of samples that was still in the dataset after rarefaction was high enough to compare the different methods, with enough samples for each method to capture the variety in its results.

Appendix 4 – Possible ecological function of abundance and mentioned genera

I wanted to investigate the potential ecological function of the most abundant genera. For this, I decided to use the 20 most abundant genera. I disregarded some more abundant ASVs, for which it had not been possible to assign them to a genus. I looked up basic ecological and physiological parameters for each of them. I investigated specifically whether they are found in symbiosis with other organisms, especially with plants or algae.

Some of the most abundant genera had previously been found in association with other eukaryotes. The type species of the genus *Agarilytica* was isolated from the algae *Gracilaria blodgettii*, the type species of the genus *Algitalea* was isolated from another algae, *Ulva pertusa*. Different members of the genus *Marinomonas* are found in different seagrass (primarily *Poseidonia oceanica*) and bivalve species [146]. The Genus *Marinoscillum* was first isolated from an unidentified sponge, and the genus *Rubidimonas* from a shrimp of the Squillidae Family (Supplementary Table 1.).

Some of the genera observed most frequently in these samples are also known for their interesting interactions with eukaryotes: Members of the genus *Sulfitobacter* are known to protect algae that they associate with from bacterial toxins [143]. Members of the genus *Aquimarina*, on the other hand, are known to be pathogenic towards crustaceans [147] (Supplementary Table 1.).

Of the most abundant genera in the sample, most were either aerobe heterotrophs, or aerobe phototrophs, mostly cyanobacteria. The genus *Methylotenera* is capable of metabolizing methylamine, an important C_1 compound in marine systems, and thus might play an important role in carbon cycling. The known members of the genus *Oleiphilus* only degrade aliphatic hydrocarbons. The genus *Celeribacter*, while normally aerobic, is capable of anaerobic nitrate reduction to nitrite, which some members of the genus *Persicirhabdus* can also do under aerobic conditions. The genus *Sulfitobacter*, finally, is capable of oxidizing sulphite and sulphur to sulphate (Supplementary Table 1.).

Supplementary Table 1. ***Changes in abundance of the most abundant genera in samples treated with different host depletion kits and their possible ecological function.*** *This table lists the 20 most abundant genera in the samples, with their relative abundances in samples extracted with PowerSoil, their relative abundances in samples depleted with QIAamp, and their possible ecological function, as well as the references for this ecological function.*

Relative abundance PowerSoil	**Relative abundance QIAamp**	**Family**	**Genus**	**Possible function/ecology**	
0.00759	0.00146	Cellvibrionaceae	*Agarilytica*	Aerobic, heterotrophic, degrades agar, isolated from the algae *Gracilaria blodgettii*	[148]
0.00304	0.01286	Flavobacteriaceae	*Algitalea*	Aerobic, metabolism unknown, isolated from the algae *Ulva pertusa*	[149]

Supplementary Table 1. ***Changes in abundance of the most abundant genera in samples treated with different host depletion kits and their possible ecological function.*** *This table lists the 20 most abundant genera in the samples, with their relative abundances in samples extracted with PowerSoil, their relative abundances in samples depleted with QIAamp, and their possible ecological function, as well as the references for this ecological function.*

Relative abundance PowerSoil	**Relative abundance QIAamp**	**Family**	**Genus**	**Possible function/ecology**	
0.01176	0.02752	Flavobacteriaceae	*Aquimarina*	Aerobic, heterotrophic, partially pathogenic in crustaceans	[147, 150, 151]
0.00418	0.02013	Arenicellaceae	*Arenicella*	Aerobic, heterotrophic	[152]
0.00711	0.012690	Pirellulaceae	*Blastopirellula*	Aerobic, phototrophic	[153]
0.00272	0.01231	Rhodobacteraceae	*Celeribacter*	Normally aerobic, heterotroph, capable of anaerobic fermenter of glucose with nitrate	[154]
0.00824	0.00728	Granulosicoccaceae	*Granulosicoccus*	Aerobic, heterotrophic,	[155]
0.00399	0.00268	Leptolyngbyaceae	*Leptolyngbya*	Aerobic, phototrophic, also in terrestrial ecosystems	[156]
0.00467	0.00546	Oceanospirillaceae	*Marinomonas*	Aerobic, partially halophilic, found also in bivalves and seagrass	[157 – 159]
0.00366	0.00690	Reichenbachiellaceae	*Marinoscillum*	Aerobic, heterotrophic, isolated from an unidentified sponge	[160]
0.00953	0.00191	Oscillatoriales Order	*MBIC10086*	Aerobic, phototrophic, filamentous	[161, 162]
0.00538	0.02026	Methylophilaceae	*Methylotenera*	Aerobic, methylamine utiliser, important in C_1 cycling in marine ecosystems	[163]

Supplementary Table 1. ***Changes in abundance of the most abundant genera in samples treated with different host depletion kits and their possible ecological function.*** *This table lists the 20 most abundant genera in the samples, with their relative abundances in samples extracted with PowerSoil, their relative abundances in samples depleted with QIAamp, and their possible ecological function, as well as the references for this ecological function.*

Relative abundance PowerSoil	**Relative abundance QIAamp**	**Family**	**Genus**	**Possible function/ecology**	
0.00266	0.00837	Oleiphilaceae	*Oleiphilus*	Aerobic, heterotrophic, only degrades hydrocarbons	[164]
0.00162	0.00635	Verrucomicrobiaceae	*Persicirhabdus*	Aerobic, heterotrophic, some species reduce nitrate to nitrite	[165]
0.03079	0.010369	Phormidesmiaceae	*Phormidesmis*	Aerobic, phototrophic	[166]
0.00325	0.00482	Lewinellaceae	*Portibacter*	Aerobic, heterotrophic	[167]
0.01286	0.06001	Verrucomicrobiaceae	*Roseibacillus*	Aerobic, heterotrophic	[165]
0.00371	0.00953	Saprospiraceae	*Rubidimonas*	Aerobic, heterotrophic, isolated from a Squillidae shrimp	[168]
0.00959	0.05830	Rubritaleaceae	*Rubritalea*	Aerobic, heterotrophic	[169]
0.01034	0.01086	Schizotrichaceae	*Schizothrix*	Aerobic, phototrophic	[170]
0.00147	0.00977	Roseobacteraceae	*Sulfitobacter*	Aerobic, heterotrophic, oxidizes sulphur and sulphite to sulphate, can protect algae from bacteria	[140, 143]
0.00425	0.01493	Roseobacteraceae	*Yoonia*	Anaerobic, heterotrophic	[171]

Based on the analysis of the fold-change of specific genera (Figure 11.), I wanted to investigate the ecological role of these microorganisms. I looked up basic ecological parameters about every genus which's abundance changed significant between the control and the host depletion with QIAamp. For those with the strongest increase, I delved deeper into their ecology. These might be the genera that are most easily overlooked with conventional sequencing. I wanted to see if some of them might potentially play important roles in seagrass ecology.

Supplementary Table 2. ***Changes in abundance of selected genera in samples treated with different host depletion kits and their possible ecological function.*** *This table lists all genera that are significantly differentially enriched in the* Zostera marina *samples between samples extracted with PowerSoil and samples depleted with QIAamp. It lists the base-2 logarithm of their fold-change between samples extracted with PowerSoil and samples depleted with QIAamp and their possible ecological function, as well as the references for this ecological function.*

Average log2 fold Change	**Family**	**Genus**	**Possible function/ecology**	
-4.84	Flammeovirgaceae	*Flexithrix*	Aerobic, heterotrophic	[172]
-3.93	Alteromonadaceae	*Glaciecola*	Aerobic, heterotrophic, psychrophilic, halophilic, isolated from sea ice, marker for healthy seagrass	[173, 174]c
-3.60	Mitochondria		Host organelle	
-3.39	Pseudobacteriovoracaceae	*Pseudobacteriovorax*	Aerobic, heterotrophic, capable of predation of other bacteria	[175]
-3.34	Cellvibrionaceae	*Agarilytica*	Aerobic, heterotrophic, degrades agar, isolated from the algae *Gracilaria blodgettii*	[148]
-3.20	Phormidesmiaceae	*Phormidesmis*	Aerobic, phototrophic	[166]
-3.09	Fulvivirgaceae	*Fulvivirga*	Aerobic, heterotrophic	[176, 177]
-3.07	Puniceicoccaceae	*Lentimonas*	Aerobic, heterotrophic	[178]
-2.98	Hyellaceae	*Pleurocapsa*	Aerobic, phototrophic, can fix nitrogen	[179, 180]
-2.97	Schizotrichaceae	*Schizothrix*	Aerobic, phototrophic	[170]
-2.95	Alteromonadaceae	*Paraglaciecola*	Aerobic, heterotrophic	[181]
-2.95	Chloroplast		Host organelle	
-2.85	Puniceicoccaceae	*Cerasicoccus*	Aerobic, heterotrophic	[182]
-2.84	Alteromonadaceae	*Salinimonas*	Aerobic, heterotrophic, halophilic	[183]
-2.60	Porticoccaceae	*C1-B045*	Aerobic, heterotrophic, degrades cycloalkanes	[184]
-2.54	Alteromonadaceae	*Aliiglaciecola*	Aerobic, heterotrophic, can reduce nitrate to nitrite	[185]
-2.52	Flavobacteriaceae	*Gilvibacter*	Aerobic, heterotrophic, can reduce nitrate to nitrite	[186]
-2.37	Puniceicoccaceae	*Pelagicoccus*	Aerobic, heterotrophic	[187]

Supplementary Table 2. ***Changes in abundance of selected genera in samples treated with different host depletion kits and their possible ecological function.*** *This table lists all genera that are significantly differentially enriched in the* Zostera marina *samples between samples extracted with PowerSoil and samples depleted with QIAamp. It lists the base-2 logarithm of their fold-change between samples extracted with PowerSoil and samples depleted with QIAamp and their possible ecological function, as well as the references for this ecological function.*

Average log2 fold Change	**Family**	**Genus**	**Possible function/ecology**	
-2.35	Lewinellaceae	*Portibacter*	Aerobic, heterotrophic	[167]
-2.30	Lewinellaceae	*Lewinella*	Aerobic, heterotrophic	[188]
-2.29	Crocinitomicaceae	*Crocinitomix*	Aerobic, heterotrophic	[189]
-2.28	Endozoicomonadaceae	*Endozoicomonas*	Aerobic, heterotrophic, can reduce nitrate to nitrite, symbiont in various marine animals	[190 – 192]
-2.28	Schleiferiaceae	*Schleiferia*	Aerobic, heterotrophic, thermophilic, reduce nitrate to elemental nitrogen	[193]
-1.92	Flavobacteriaceae	*Zobellia*	Aerobic, heterotrophic, isolated from red algae, can utilise complex algal carbohydrates	[194]
-1.77	Marinicellaceae	*Marinicella*	Aerobic, heterotrophic	[195]
-1.68	Granulosicoccaceae	*Granulosicoccus*	Aerobic, heterotrophic,	[155]
-1.66	Colwelliaceae	*Colwellia*	Facultative anaerobic, heterotrophic, obligate barophilic, psychrophile	[196, 197]
-1.50	Roseobacteraceae	*Yoonia*	Anaerobic, heterotrophic	[171]
-1.39	Cellvibrionaceae	*Eionea*	Aerobic, heterotrophic	[198]
-1.14	Pirellulaceae	*Blastopirellula*	Aerobic, phototrophic	[153]
-0.24	Flavobacteriaceae	*Ulvibacter*	Aerobic, heterotrophic, isolated from the green algae *Ulva fenestrata*	[199]
0.75	Rhodobacteraceae	*Celeribacter*	Normally aerobic, heterotroph, capable of anaerobic fermentation of glucose with nitrate	[154]
1.15	Rhizobiaceae	*Lentilitoribacter*	Aerobic, heterotrophic	[200]
1.17	Rhodobacteraceae	*Boseongicola*	Aerobic, heterotrophic	[201]
1.25	Flavobacteriaceae	*Olleya*	Aerobic, heterotrophic	[202]

Supplementary Table 2. ***Changes in abundance of selected genera in samples treated with different host depletion kits and their possible ecological function.*** *This table lists all genera that are significantly differentially enriched in the* Zostera marina *samples between samples extracted with PowerSoil and samples depleted with QIAamp. It lists the base-2 logarithm of their fold-change between samples extracted with PowerSoil and samples depleted with QIAamp and their possible ecological function, as well as the references for this ecological function.*

Average log2 fold Change	Family	Genus	Possible function/ecology	
1.36	Methylophilaceae	*Methylotenera*	Aerobic, methylamine utiliser, important in C_1 cycling in marine ecosystems	[163]
1.42	Pirellulaceae	*Rhodopirellula*	Aerobic, heterotrophic, isolated from macroalgae	[153, 203, 204]
1.51	Roseobacteraceae	*Octadecabacter*	Aerobic, heterotrophic, psychrophilic, isolated from sea ice and marine red algae	[205, 206]
1.56	Ahrensiaceae	*Pseudahrensia*	Aerobic, heterotrophic, can reduce nitrate to nitrite	[207]
1.62	Hyphomonadaceae	*Litorimonas*	Aerobic, heterotrophic, halophilic, some species isolated from the algae *Cladophora stimpsoni*	[208, 209]
1.80	Roseobacteraceae	*Pelagicola*	Aerobic, heterotrophic	[210]
1.81	Verrucomicrobiaceae	*Roseibacillus*	Aerobic, heterotrophic	[165]
1.87	Arenicellaceae	*Arenicella*	Aerobic, heterotrophic	[152]
1.90	Arcobacteraceae	*Halarcobacter*	Aerobic, can grow anaerobically, heterotrophic, reduces nitrate to nitrite	[211]
1.90	Saprospiraceae	*Rubidimonas*	Aerobic, heterotrophic, isolated from a Squillidae shrimp	[168]
2.00	Rhodobacteraceae	*Amylibacter*	Aerobic, heterotrophic	[212]
2.07	Rubritaleaceae	*Rubritalea*	Aerobic, heterotrophic	[169]
2.08	Roseobacteraceae	*Litoreibacter*	Aerobic, heterotrophic, psychrotolerant, some species isolated from the sea squirt *Halocynthia roretzi*	[213, 214]

Supplementary Table 2. ***Changes in abundance of selected genera in samples treated with different host depletion kits and their possible ecological function.*** *This table lists all genera that are significantly differentially enriched in the* Zostera marina *samples between samples extracted with PowerSoil and samples depleted with QIAamp. It lists the base-2 logarithm of their fold-change between samples extracted with PowerSoil and samples depleted with QIAamp and their possible ecological function, as well as the references for this ecological function.*

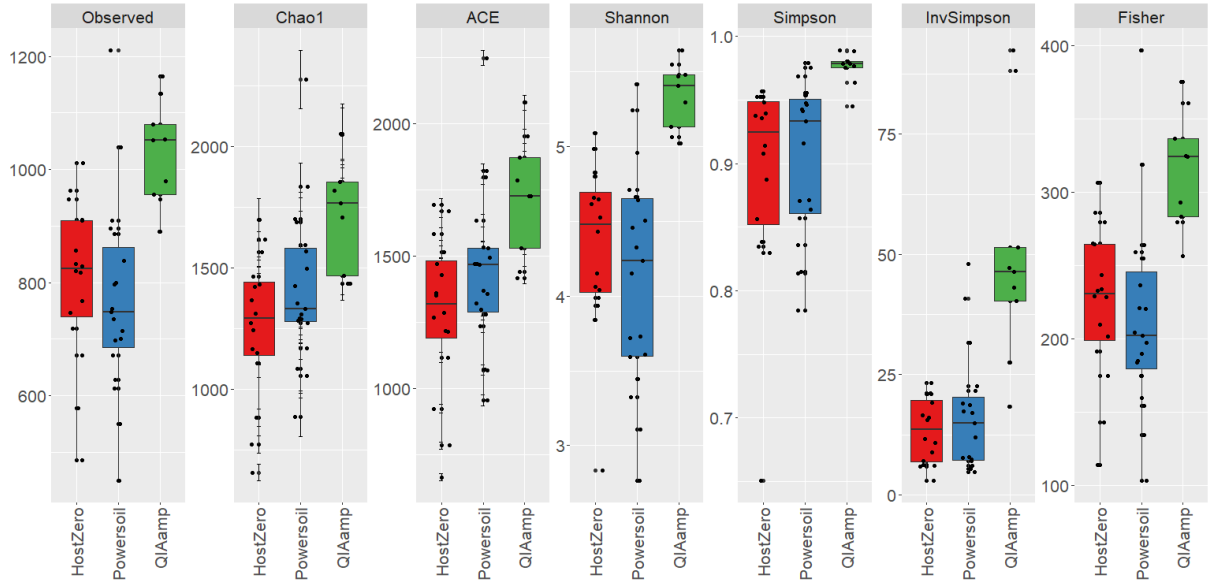
Average log2 fold Change	**Family**	**Genus**	**Possible function/ecology**	
2.25	Roseobacteraceae	*Planktotalea*	Aerobic, heterotrophic, some members are capable of anoxygenic photosynthesis	[215]
2.42	Flavobacteriaceae	*Aurantivirga*	Aerobic, heterotrophic	[216]
2.43	Cryomorphaceae	*Vicingus*	Aerobic, heterotrophic, mesophilic	[217]
2.65	Roseobacteraceae	*Sulfitobacter*	Aerobic, heterotrophic, oxidizes sulphur and sulphite to sulphate, can protect algae from bacteria	[140, 143]
2.80	Hyphomonadaceae	*Robiginitomaculum*	Aerobic, heterotrophic, can reduce nitrate to nitrite	[136]
2.80	Flavobacteriaceae	*Polaribacter*	Aerobic, heterotrophic, psychrophilic, can utilise complex algal carbohydrates, might provide them to seagrass	[124, 132]

Appendix 5 – Supplementary results

α-diversity behaved the same across all different indices

I wanted to investigate more indices of α-diversity to see if the trend described above holds true for more of them. I performed the same analyses as before, but for a broader range of indices. I included, in addition to the indices described above, the ACE index, the Simpson and the inverse Simpson index of diversity, and the Fisher index of diversity. For all of them, I did statistical analysis.

A very similar pattern than with the indices described above was observed with both the ACE indices, another index that aims to calculate the actual richness from the observed one. The main difference to the other indices here was that the median for the samples depleted with HostZero is the lowest, and that the difference between the samples treated with QIAamp and all the others was less pronounced than in the observed richness (Supplementary Figure 4.).



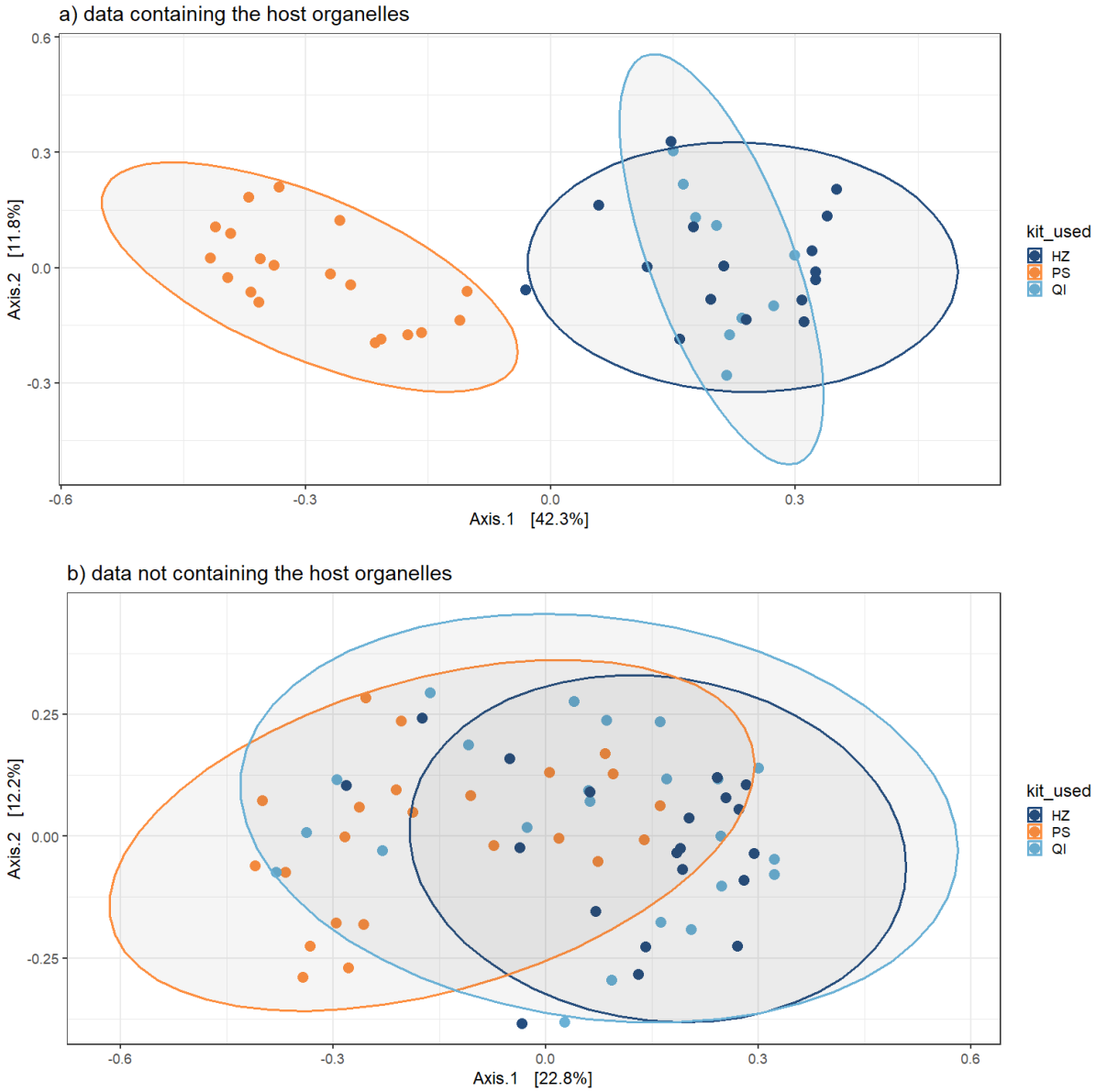
Supplementary Figure 4. ***α-diversity of samples treated with different host depletion kits.*** *The y-axis depicts the value of the α-diversity index shown in the title of each sub-panel (from left to right observed diversity, Chao1, ACE diversity, Shannon diversity, Simpson diversity, inverse Simpson diversity, and Fisher diversity). The x-axis lists the method that was used for host depletion. The boxplots cover the area from the 25th to the 75th percentile, while the arms cover the rest of the samples, excluding outliers. Each point represents one single sample. The observed diversity is not increasing significantly HostZero (p=0.5754), but it is increasing significantly for QIAamp (1.932x10^{-5}). The Chao1 index is not increasing significantly for HostZero (p=0.1901), but it is increasing significantly for QIAamp (p=4.294x10^{-3}). The ACE index is not increasing significantly for HostZero (p=0.1261), but it is increasing significantly for QIAamp (p=0.01218). The same pattern is true for the Shannon index, it is not increasing significantly for HostZero (0.3543), but it is for QIAamp (8.75x10^{-7}). The same is also true for the Simpson index (HostZero p=0.6724, QIAamp p=9.13x10^{-5}), the Inverse Simpson index (HostZero p=0.328, QIAamp p=0.003123), and the Fisher index of diversity (HostZero p=0.5731, QIAamp p=3.999x10^{-5}).*

For the Simpson index, the QIAamp sample still had the highest value. In this case, though, this means something else, because for the Simpson index, the lower the index, the higher the diversity. Nevertheless, the samples depleted with the QIAamp kit had the highest Simpson index at around 0.975, with the smallest range, from 0.945 to 0.99. The other sets of samples, on the other hand, had a far larger range, ranging from 0.775 to 0.975, with their medians being all at around 0.92, with only the one for the samples just extracted with the PowerSoil kit slightly higher (Supplementary Figure 4.).

For the Inverse Simpson, a higher index again means a higher diversity, and again, the sample depleted with QIAamp had the highest value with a median of about 45, while the others were all between 12 and 15. The strongest outliers in this case were three samples depleted with the QIAamp kit: All three

of them had an Inverse Simpson index between 87 and 95. The Fisher index exactly followed the same pattern as the observed richness. The samples depleted with QIAamp had the highest index at about 325, while the median of the other three was at about 200, except for the samples depleted with HostZero, where it was slightly higher, at about 230 (Supplementary Figure 4.).

All β-diversity analyses showed similar results

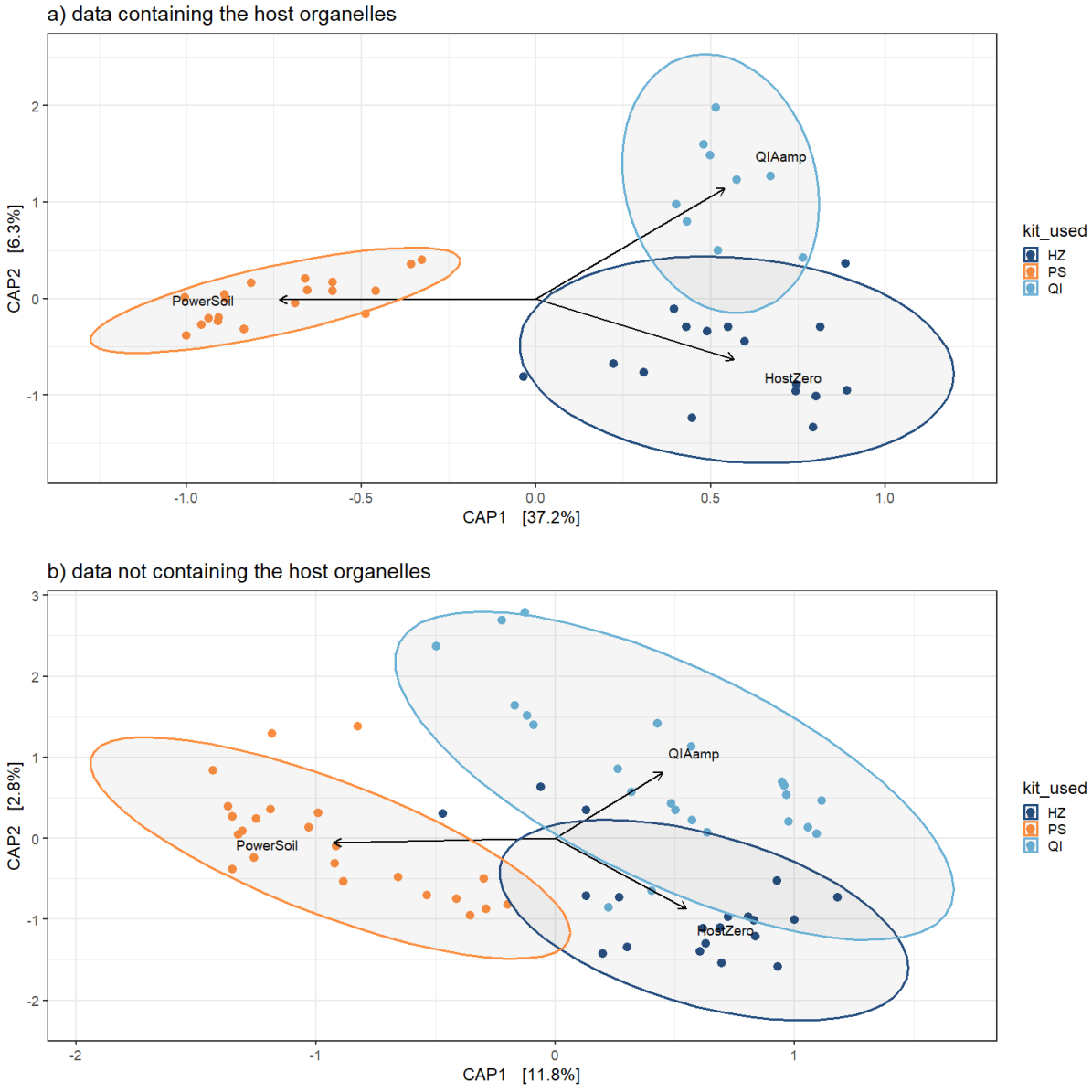


Supplementary Figure 5. ***PCoA analysis of community composition of samples treated with different host depletion kits.*** *The two axes depict the two primary components. Each point represents a single sample, with their distances in the plot representing the real distance of their community composition. Different colours represent different treatments. The ellipses represent the 95% confidence interval of each treatment. a) depicts the differences in community composition as they are. All samples are changing significantly with host depletion. The change is clearly significant for HostZero ($p>10^{-4}$, $R^2=0.431$) and QIAamp ($p>10^{-4}$, $R^2=0.403$). b) depicts the differences in community composition as they are, but with the host organelles removed* in-silico. *HostZero ($p=3x10^{-4}$, $R^2=0.158$) and QIAamp ($p=3x10^{-4}$, $R^2=0.129$) are changing the community significantly.*

I wanted to investigate more representations of the of β-diversity of the data to see if the same patterns were still present. I performed both principal coordinate analysis, PCoA, and constrained analysis of principal coordinates, CAP. I performed both analyses two times, like with the NMDS analysis. I first did

it with the whole sequenced data included. Then, I did with the dataset where the host organelles had been removed *in-silico*, to see the changes that were happening in the microbial community. I also analysed all these plots statistically.

In the PCoA, there was a clear difference between host depleted samples and the control. The first axis explained a large part of the variation, 42.3%. It was a clear cut between those two sets of methods. There was a small overlap between the confidence interval of the samples depleted with the two host depletion kits, but there were no samples that fell into this overlap. This difference between the samples and the control was statistically significant ($p>10^{-4}$ for both host depletion kits; Supplementary Figure 5.).



Supplementary Figure 6. ***CAP analysis of community composition of samples treated with different host depletion kits.*** *The two axes depict the two primary components. Each point represents a single sample, with their distances in the plot representing the real distance of their community composition. Different colours represent different treatments. The ellipses represent the 95% confidence interval of each treatment. a) depicts the differences in community composition as they are. All samples are changing significantly with host depletion. The change is clearly significant for HostZero ($p>10^{-4}$, $R^2=0.431$) and QIAamp ($p>10^{-4}$, $R^2=0.403$). b) depicts the differences in community composition as they are, but with the host organelles removed* in-silico. *HostZero ($p=3x10^{-4}$, $R^2=0.158$) and QIAamp ($p=3x10^{-4}$, $R^2=0.129$) are changing the community significantly.*

When removing the host organelles from the analysis, the two sets of data moved closer together, and some overlap was introduced between the PowerSoil and HostZero kits. In addition, none of the two main components fully split the data anymore, and they also lost a lot of their explanatory power, going down to just 22.8% for the first axis. The differences between the two kits also became less pronounced. There were now some HostZero samples that fell inside of the confidence interval of the QIAamp kit, and vice versa. The difference between the host depletion and the control was still statistically significant ($p=3x10^{-4}$ for both samples). At the same time, the explanatory power of the difference in treatment decreased, from an R^2 of 0.4 to one at about 0.15 (Supplementary Figure 5.).

In the constrained analysis of principal coordinates, the first axis explained 37.2% of variation (Supplementary Figure 6.), which was slightly lower than in the PCoA (Supplementary Figure 5.). There was again the same observable split described for the other two analyses along this axis between host depleted samples and the control. This split was statistically significant ($p>10^{-4}$ for both host depletion kits). Another difference was that the second axis of the CAP plot explained only 6.3% of variation (Supplementary Figure 6.), not 11.8% like in the PCoA (Supplementary Figure 5.).

When removing the host organelles from the analysis, the same happened as in the principal component analyses: The first axis lost most of its explanatory power, going down to just 11.8% of the present variation. The second axis also lost explanatory power, going down to just 2.8%. But the differences were still significant ($p=3x10^{-4}$ for both samples), but with the same loss of explanatory power described before. The R^2 decreased from about 0.4 to about 0.15 (Supplementary Figure 6.).