universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Microbiome of the Photosymbiotic Waminoa Flatworm: Method Development and Characterization

verfasst von | submitted by

Simone Vanessa Müller BSc

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

Master of Science (MSc)

Wien | Vienna, 2025

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:	Elizabeth Hambleton BA PhD

Abstract

Waminoa is a genus of marine acoel flatworm that harbors unicellular dinoflagellate symbionts which supply it with products of photosynthesis, a relationship called photosynthetic symbiosis, or ‘photosymbiosis’. The importance of the microbial community present in a photosymbiotic system, beyond the photosynthetic symbiont, has been increasingly recognised in the past years due to its nutrient-based as well as defensive symbiosis that it forms with the host, and its potential interaction with the photosymbiotic relationship present in the system. While the algae symbiont has been characterized with genetic markers, there has been no characterization of the bacterial and archaeal microbial community of *Waminoa*. Characterization is initially achieved by high-throughput amplicon sequencing of the 16S small subunit ribosomal marker, which is a commonly used genetic marker conserved across taxa. However, amplicon sequencing of the bacterial 16S DNA is complicated due to the high amount of host DNA overwhelming samples, making it difficult to access bacterial DNA in amplicon sequencing. In this thesis, I have tested multiple methods to remove host DNA from whole tissue extracts and sequenced the attained bacterial DNA. The Qiagen QIAamp DNA Microbiome kit performed the best in terms of host depletion. Development of Locked Nucleic Acid (LNA)-based PCR clamps, designed to block amplification of host DNA, failed to effectively achieve host depletion. I discuss potential bias produced by different DNA extraction and host depletion methods. Finally, I show comparisons of the microbial communities of two lines of symbiotic worms and one line of aposymbiotic worm. However, none of these are statistically significant, and it is not clear if this is due to biological reality of a low abundance / low diversity community, or technical issues resulting in insufficient detection of the microbiome. Further testing and optimizing of bacterial DNA extraction and host DNA depletion is needed to answer this question.

Deutsche Zusammenfassung

Waminoa ist eine Gattung mariner acoelomater Plattwürmer, die von in photosynthetischer Symbiose oder “Photosymbiose” mit ihnen lebenden einzelligen Dinoflagellaten mit Produkten derer Photosynthese versorgt werden. Die Bedeutung des Mikrobioms in photosymbiotischen Systemen, über den photosynthetischen Symbionten hinaus, wurde in den vergangenen Jahren zunehmend anerkannt, aufgurnd der Nährstoff-basierten und defensiven Symbiosen, die Mikroben mit dem Wirtsorganismus eingehen, als auch wegen dem Potenzial der Mikroben, mit der photosymbiotischen Interaktion im System zu interagieren. Während die symbiotische Alge mit genetischen Markern bereits charakterisiert wurde, gibt es noch keine Beschreibung der Bakterien und Archaea im Mikrobiom von *Waminoa*. Eine Beschreibung wird zuerst durch High-Throughput Amplicon Sequencing des 16S Small Subunit Markers erstellt, welches ein häufig verwendetes Marker-Gen ist, das über Taxa hinweg konserviert ist. Der hohe Anteil an der Wirts-DNA kann jedoch die bakterielle DNA überwältigen und Schwierigkeiten bereiten, bakterielle DNA durch Amplicon Sequencing zu erreichen. Im Lauf dieser Masterarbeit habe ich mehrere Methoden getestet, um Wirts-DNA von Gewebe-Proben zu entfernen und die dadurch erhaltene bakterielle DNA zu sequenzieren. Das Qiagen QIAamp DNA Microbiome kit entfernte Wirts-DNA am effektivsten. Die Entwicklung von PCR Clamps auf Locked Nucleic Acid (LNA) Basis, zur Vermeidung von Vermehrung der Wirts-DNA während der PCR, erreichte nicht den erwarteten Effekt. In meiner Diskussion gehe ich auf den potenziellen Einfluss der verschiedenen Methoden auf Endergebnisse ein. Zuletzt vergleiche ich die Mikrobiome zwei verschiedener symbiotischer Wurmlinien und einer aposymbiotischen Wurmlinie. Diese Unterschiede sind jedoch nicht statistisch signifikant, und es ist unklar, ob dies der biologischen Realität eines Mikrobioms mit niedrigem Artenreichtum zugrunde liegt, oder technischer Schwierigkeiten folgend einer nicht ausreichenden Erfassung des Mikrobioms. Fortsetzung der Tests und der Verbesserung der Extraktion bakterieller DNA sowie Verbesserung der Entfernung der Wirts-DNA sind notwendig, um diese Frage zu beantworten.

Acknowledgements

First of all, I would like to thank Elizabeth (Liz) Hambleton, my group leader and supervisor. She has made this thesis possible by giving me the opportunity to join her research group and taking valuable time to support me and my project.

I would also like to thank Sandra Montaño Salazar for her mentorship and her input, sacrificing time out of her own demanding projects to assist me with mine. I must acknowledge her contributions to this thesis, without which it could not have happened.

My thanks also go to animal care technician Cátia Pinto, whose reliable work and innovations were vital to my project. Her work is a fundamental part of the research in this group.

The final group member I would like to thank personally is Marlis Stenzl, my fellow master's student, who supported me on both an intellectual level and a personal level. To learn together and to struggle together shapes a unique bond.

Furthermore, I would like to thank the Joint Microbiome Facility, especially Petra Pjevac, Joana Silva, Gudrun Kohl, and Jasmin Schwarz, for their continued support throughout the project. In addition to performing all the sequencing, they have taken extra efforts for me to be able to try new methods as part of this thesis.

Working at DOME has been a great opportunity, and one that I have been interested in for a long time. It would not have been possible without the reliable work of the DOME administration team, and the DOME technician team, whom I would like to thank for their support in day to day research life.

Finally, I would like to thank Daan Speth for his valuable advice in the hallways, my fellow master's students for creating a truly supportive and friendly working environment, and my partner and my family for supporting me during stressful stretches of my work.

Table of Contents

1. Introduction

1.1. The acoel flatworm *Waminoa*

Waminoa is a genus of marine acoel flatworms living on the surface of various soft and stony corals[1] at a depth of 5m or more[2] where it feeds on coral mucus, but not tissue[3]. It is a flat, disc shaped worm that appears bronze colored due to dinoflagellate endosymbionts[1], identified as belonging to the genera *Symbiodiniaceae*[4] and *Amphidinium*[3]. Its ability to host and vertically transmit two different symbiotic algae at the same time, in some cases within the same cell, is one of its notable features[5]. In captivity, *Waminoa* can be kept isolated from its coral host in small tanks of seawater in an incubator. This, and a sufficiently short generation time, makes it viable as a model organism to study photosymbiosis. Some of its advantageous characteristics as a model organism are its ability to reproduce clonally in captivity, which produces lines of genetically identical animals, and a process in which symbiotic algae can be removed from the animal without killing the host, creating 'aposymbiotic' individuals. This has been used to produce sets of genetically highly similar lines of symbiotic and aposymbiotic *Waminoa* worms in the facilities of the Division of Microbial Ecology of the University of Vienna. It mainly serves research interests in the Hambleton group pertaining to the symbiosis between animal and algae, where non-coral systems are currently underrepresented.

1.2. Symbiosis in marine invertebrates

Symbiosis is defined as the "living together of unlike organisms" by Heinrich Anton de Bary[6,7]. In its current usage, the term refers in most sources to a relationship that is both intimate and constant, and if it is mutually beneficial, it is referred to as a „mutualistic symbiosis" or „mutualism"[7]. Photosymbiosis, in which a symbiont capable of photosynthesis shares products with its host[8], is an essential mutualism in coral reefs[9]. "Coral bleaching", the loss of algae from the coral due to stress[10], has become an increasingly pressing issue in the marine environment[11,12]. It has contributed to the loss of about 14% of the coral from the world's coral reefs between 2009 and 2018[13]. In *Waminoa*, the animal hosts unicellular algae belonging to *Symbiodiniaceae*[4] and *Amphidinium* that are genetically distinct from their resident coral's algal symbionts, meaning they receive their symbionts from another source[3]. The symbiosis between *Waminoa* and its algae is believed to center around nutrients such as sterols produced by the algae, though the full scope of this symbiosis is still being investigated[14].

The exchange with algae is, however, not the only relationship harbored by the animal host: marine animals live alongside a vast variety of microorganisms, including bacteria, archaea, protists, and fungi, and many of these microorganisms live in close symbiosis with marine animals as a microbiome[15]. This includes, of course, the aforementioned corals, but a great variety of bacteria are known to live within a vast range of marine invertebrate systems. In a 2021 study, over 400 biosynthetic gene clusters for secondary metabolites were found in 74 cultured strains of coral bacterial symbionts, highlighting the breadth of the coral host's symbionts alone[16]. Some functions that bacterial symbionts may provide are nutrient based, such as the fixation of nitrogen, or defensive, such as the production of antibiotics[17]. Conversely, multiple studies have found *Endozoicomonas* in a variety of marine invertebrate hosts, including sponges, molluscs, and worms, as well as photosymbiotic systems such as corals[18]. Its role in these systems is not fully understood, though it may provide diverse functions by adapting to different hosts, which could include the upcycling of carbohydrates or supplying protein[18].

This large array of interactions opens new doors for hosts: changes in symbiotic composition can outpace sexual generation and provide a quicker method of adaptation[19,20]. Given the accelerated loss of coral reefs in recent years, this has become a particular talking point surrounding corals. Peixoto *et al*. suggest that the bacterial community of a coral has an influence on the coral's heat resistance and bleaching events, and proposes a system of identifying beneficial bacterial symbionts for corals (Beneficial Microorganisms for Corals, BMC)[21]. Other photosymbiotic cnidarians have been investigated as well: in the freshwater cnidarian *Hydra*, significant differences in bacterial communities were reported for symbiotic and aposymbiotic genetically identical animals, and transfer of *Legionella* from aposymbiotic animals to symbiotic animals caused the release of algae symbionts and the subsequent uptake thereof by aposymbiotic animals[22].

For the bacterial communities of *Waminoa*, we can therefore form the hypothesis that the presence or absence of algal symbionts in the system affects the bacterial community. The removal of algae symbionts may open niches that they previously occupied, making room for taxa to enter or reach higher abundances as compared to symbiotic animals. The algae's involvement in the worm's defenses is currently under investigation, but its absence may open doors for pathogens that the algae could guard against. Overall, I expect the richness of the system to decrease similarly as in the previously mentioned study on *Hydra*[22], and I expect an increase of pathogens, due to potential loss of defense mechanisms. To test this, the bacterial community needs to be identified through sequencing of conserved genetic markers, but this introduces challenges that must be solved first.

1.3. Sequencing the microbial community of a system

Amplicon sequencing is a method for studying bacterial communities, in which a selected gene is amplified through polymerase chain reaction (PCR) to make it more available for sequencing[23,24]. Amplicon sequencing of the 16S small subunit ribosomal RNA (rRNA) gene is considered one of the most common and effective ways of characterizing a bacterial community[25], and has found application in many large-scale projects over the last decade. This includes projects investigating the symbiotic communities residing in a macroorganism host, such as the Human Microbiome Project (HMP)[26]. The traditional method of culturing bacteria for identification is estimated to only cover 1% of the diversity present in an environmental sample[27], so the introduction of community-based high-throughput sequencing methods has revolutionized the field.

Amplicon sequencing is not without flaws, however. A wide variety of available methods has led to a lack of standardized protocol, best practices, and properly comparable and replicable data[23]. The DNA extraction method has been shown to influence the resulting community composition[28]. The choice of which 16S hypervariable region to be targeted, for instance, can affect outcomes, and different regions are more or less suitable for different samples or research questions[25]. Effects of PCR, such as choice of primer[29,30] and cycle number[30] have also been shown to impact results to such a degree that comparisons across different protocols cannot be reliably made. Choice of bioinformatics software is yet another factor affecting results[23].

Extraction and sequencing of bacterial DNA from a eukaryotic host poses a unique challenge in the removal of the host DNA from the sample. In amplicon sequencing, although eukaryotic hosts do not possess the 16S region, cross-binding by primers can lead to erroneous amplification of the eukaryotic 18S rRNA region. Depending on the site, samples can contain > 90% host-aligned reads in human samples[31], and host DNA can overwhelm a sample and disturb microbial signal, causing problems in downstream analysis[32]. A variety of methods have been developed to reduce host signal, but these methods may introduce additional bias in the resulting data[33]. These methods can roughly be categorized into physical methods such as filtering, DNA extraction methods, and PCR based methods.

1.3.1. Physical methods

The separation of host and symbiont DNA before extraction is not a trivial matter, although it has been done using filtration on sputum and urine[34,35]. When considering symbionts that might be intracellular, filtration however runs the risk of removing these symbionts along with host cells[36]. Not only is this to be expected in an animal host, but cited studies have used filtration on bodily

fluids and not whole animals, which need to be processed (often by homogenization) into a liquid form to be able to pass a filter. It is unclear if this step could introduce additional problems.

Centrifuging could also be used to separate host cells from bacterial cells. Another study on saliva however found no depletion of human DNA by centrifugation[31]. When it comes to animal tissue rather than fluids, similar problems could arise as in filtration methods. The HostZERO Microbial DNA kit from Zymo Research (discussed below) nonetheless includes a centrifugation step in combination with biochemical lysis, which can circumvent arising problems, but is not guaranteed to be effective. In early testing, I saw no effective results using filtration and centrifugation without biochemical preparation beforehand. For this reason, I have decided to focus on biochemical methods.

1.3.2. Host depletion during DNA extraction

The method of DNA extraction affects study results, and this has been known for over a decade[37,38]. This is also true for methods of host depletion: for instance, some methods have been shown to introduce bias against specific types of microbes in bovine milk[33]. Nonetheless, the increase in proportion of bacterial DNA in the total DNA from a reported under 10% to over 70%[32] has an undeniable impact for samples overwhelmed with host DNA.

Commercial kits are available for DNA extraction with host depletion steps. In a study by Heravi *et al*[32], the QIAamp DNA Microbiome kit from Qiagen showed a 32-fold improvement from the negative control in 18S/16S ratio, and significantly depleted host reads in a separate experiment on saliva[31]. The HostZERO Microbial DNA kit produced by Zymo Research saw promising results in two studies[32,39]. In a study on bovine and human milk[40], out of three methods tested, the MolYsis complete5 kit from Molzym performed best. Other available kits are, for example, the Molzym Ultra-Deep MicrobiomePrep kit and the NEBNext® Microbiome DNA Enrichment kit, which were not tested in this thesis due to time and financial constraints. Most of the available kits follow the lysis approach as described below and differ primarily by manufacturer and proprietary buffer composition, with the exception of the NEBNext® Microbiome DNA Enrichment kit, which makes use of the increased methylation of human DNA CpG sites to remove human DNA via a binding protein.

The common approach in host depletion kits and protocols is to selectively lyse host cells and degrade the released host DNA while keeping bacterial cells intact, and then perform a total DNA extraction[36]. There also exist protocols that are for use without kits, for example, a method

described by Marotz *et al.*[31] in 2018 called 'lyPMA': this method uses propidium-monoazide, and a method described by Charalampous *et al.*[41] based on saponin.

1.3.3. Host depletion during PCR

An alternative to host depletion during extraction is to prevent host DNA from being amplified during PCR through chemical blockers. These come mainly in two forms: Locked Nucleic Acid (LNA) and Peptide Nucleic Acid (PNA)[42], though other options like blocking primers[43] are available. These molecules bind tightly to template DNA and can either prohibit the PCR primer from binding, or block elongation[42].

LNAs are nucleotides that are more stable than regular nucleotides because of a methylene bridge in the deoxyribose sugar[42] and have found application in a variety of use cases, including PCR primers[44], antisense oligonucleotides[45], and of course PCR blockers[42]. It can be synthesized similarly as conventional oligonucleotides, and has a relatively low cost[46], which makes it a viable option for custom designed host depletion. Guides for designing custom LNAs exist[42,46], but they are not numerous and best practices are still in development.

PNAs on the other hand have an amide linked backbone in place of the sugar phosphate backbone of DNA[47]. They also selectively block amplification, they bind more strongly and they are not susceptible to nucleases; however, they are also significantly more expensive and take longer to synthesize[42].

1.4. Aim and scope of this thesis

For this thesis, my aim was to test and develop methods for high throughput amplicon sequencing of the bacterial community of the *Waminoa* flatworm. These methods, once developed, could be repeatedly employed to better understand *Waminoa* and make it more available as a model system, as well as help develop similar methods for other marine invertebrate systems, especially other acoels, and other photosymbiotic non-coral organisms. For this purpose, I have sequenced the bacterial community of the *Waminoa* flatworm using multiple methods and compared the resulting communities. I have included the QIAamp DNA Microbiome kit for its special emphasis on host DNA removal and microbial DNA enrichment, and a selection of different LNAs to examine the potential of PCR blocking. I tested, but did not sequence, an additional array of DNA extraction methods, such as filtering, centrifuging, and the Zymo HostZERO Microbial DNA kit. As comparisons, I used the DNeasy PowerSoil Pro kit and the DNeasy PowerBiofilm kit from Qiagen as kits that do not have specialized steps to eliminate host DNA. I chose to omit non-kit methods

from this thesis due to time constraints as well as the demonstrated success in published studies of the aforementioned kits, since it is to be expected that non-kit methods require multiple test runs to optimize.

2. Materials & Methods

2.1. Animals

For this study, I used adult *Waminoa* worms kept by the Hambleton group in the marine culturing facility of the University of Vienna Biology Building (UBB), Vienna. Worms were kept separated by clonal line. The lines I used were clonal lines “C” and “A”, which were collected from the Haus des Meeres through collaboration with Daniel Abed-Navandi in 2021.

The animals were kept in glass tanks (Ikea 365+ Food Containers, 600 ml, square, silicon lid) in an Percival Plant Climatics incubator at 25°C on a 10:14 light:dark cycle with a light level of 12 mol photons $m^{-2}s^{-1}$. Tank changes and seawater changes were each done weekly. Filtered artificial seawater (FASW) was made by dissolving sea salt (Tropic Marine Reef Salt) in Reverse Osmosis water for 48h under agitation to a concentration of 31.5-32.5 ppt, verified with a salinity meter, then filtered through a 0.2 µm filter into sterile containers and stored at 18°C. Animals were fed once per week with freshly hatched nauplii of *Artemia salina* brine shrimp.

Aposymbiotic animals (animals without symbiotic algae) were provided by lab technician Cátia Pinto. They had been created using an in-house protocol developed by Cátia Pinto prior to the beginning of my experiment, in which symbiotic line C worms were treated with a cycle of exposure to menthol for two short periods within a week, with intermittent rest periods, for four weeks. Animals were controlled for any remaining symbiotic algae via visual inspection of algae autofluorescence with fluorescent microscopy. After treatment, aposymbiotic worms were kept under the same conditions as the symbiotic worms described above.

2.2. Initial testing of DNA extraction methods

To select DNA extraction methods prior to sequencing, I tested several commercial kits and physical separation methods (for all extraction information, see Appendix Table 1). In all these tests, animals were starved for at least 3 days, to allow full digestion and thus avoid contamination from food particles. Per test sample, I took 5-15 symbiotic C-line worms and rinsed them at least three times with FASW to remove any loosely attached external microbes. I then transferred the animals into 1.5ml reaction tubes and removed as much FASW as possible by pipetting before proceeding with extraction.

Kits I tested were the Zymo HostZERO Microbial DNA kit (Zymo Research), ZymoBIOMICS DNA Miniprep kit (Zymo Research), and the Qiagen DNeasy PowerSoil Pro kit with an altered

protocol (lysis step extended for 24h). In the case of the ZymoBIOMICS DNA Miniprep kit, I used a maceration tool (IKA T-10 Basic Ultra-Turrax homogenizer with S10N-5G steel tip) to homogenize the animals. For the Zymo HostZERO Microbial DNA kit, worms were homogenized by adding MilliQ water instead, which between tests had been found to be a faster method. After this, I followed manufacturer's instructions unless specified otherwise.

Sandra Montaño Salazar, a PhD student in our group, kindly assisted me in testing. She tested the Qiagen DNeasy Blood and Tissue kit with an altered protocol (lysis step extended overnight in 56°C waterbath) and otherwise as per manufacturer's instructions.

For testing physical separation methods, I macerated animals as above and then attempted to separate bacteria prior to extraction using centrifugation and/or filtration. The centrifugation settings I tested were: 3 min at 4300 x g, 2 min at 24 x g, 1 min at 94 x g, 2 min at 94 x g. I tested syringe filtering by removing the plunger of a sterile 1000µl syringe with a 5µm filter attached, then pipetting the sample in from the top, reattaching the plunger, and plunging down as much as was possible. I also tested centrifugation filters which were attached on top of a 1.5ml tube, then sample was added into the filter, and the filter and tube combination was centrifuged at 100 x g for 3 min.

The resulting DNA yield from extractions was measured either with a NanoDrop spectrophotometer (Model ND-1000) or Qubit fluorometer (Model Qubit 4, Invitrogen). The above tests yielded either poor to no yields (all samples where separation of host and bacteria cells was attempted using filtering or centrifuging, Zymo HostZERO Microbial DNA kit) or showed little difference (ZymoBIOMICS DNA Miniprep kit without additional treatment, QIAGEN DNeasy Blood and Tissue kit) to the kits described below that were used for subsequent sequencing. Detailed information is listed in Appendix Table 1.

2.3. DNA extraction

To extract total DNA for subsequent sequencing, I used the Qiagen DNeasy PowerBiofilm kit, the Qiagen DNeasy PowerSoil Pro kit, as well as the QIAamp DNA Microbiome kit. Extractions using the first two, the DNeasy PowerBiofilm and DNeasy Powersoil, were performed in one round on the same day (Appendix Table 1). For these, I used ten worms per sample and three samples per worm type (symbiotic and aposymbiotic C-line for PowerBiofilm; symbiotic and aposymbiotic C-line as well as symbiotic A-line for PowerSoil). As above, animals were starved for at least 3 days to avoid contamination from food particles. Animals were rinsed with FASW five times to dislodge loosely associated external microbes, placed in kit-provided tubes, and as much FASW as possible was removed by pipetting. Samples were centrifuged in the first step of the kit, which killed the

animals immediately, meaning no separate killing step was necessary. I did however introduce a separate killing step when using the QIAamp DNA Microbiome kit by adding MilliQ water to the animals, which killed them immediately. For the seawater control sample, I filtered 300 ml of tank water through a 0.2 µm polycarbonate filter (Millepore Isopore PC membrane, GTTP02500) filter using a sterile syringe and filter holder. I then removed the filter with sterilized tweezers and placed it in a kit-provided tube. After this, I followed the manufacturer's instructions. For the QIAamp DNA Microbiome kit, which in contrast to the other two kits has steps to deplete host DNA, I performed one round of extractions using ten worms per sample and three samples per worm type (symbiotic and aposymbiotic C-line, and symbiotic A-line). As above, worms were collected shortly before feeding. I rinsed them three times with FASW, placed them in 1.5ml tubes, and removed as much FASW as possible. I replaced the fluid with 900µl of MilliQ water, which kills the animals immediately and dissolves the tissue, also making any further homogenization unnecessary. For the seawater control sample, I filtered 150 ml of surrounding seawater from each animal tank as described above. I then proceeded following manufacturer's instructions.

For all three kits, I also generated extraction 'blanks', in which an empty 1.5 ml tube was the starting step, and then the extraction proceeded identically to the other samples. All extracted DNA was quantified by NanoDrop or Qubit as described above (see Appendix Table 1) and then stored at -20°C until library preparation for sequencing.

2.4. Bacterial isolates

As a comparison to the bacterial community identified from sequencing, we also analyzed bacterial isolates that were generated and kindly provided by Sandra Montaño Salazar, a PhD student in our group. For this, 10 symbiotic C-line worms were macerated, serial dilutions with filtered artificial sea water were performed and the dilutions 10^2, 10^5 and 10^7, were used to cultivate bacteria on marine agar plates (Difco 2216 Marine Broth, agar 1.5%), grown at 25°C during 3 days. Bacterial colonies were streaked onto the plate surface using an inoculating loop. Once the colony was pure, one bacterial colony of each culture was added into a PCR tube with Milli-Q water in order to perform a colony PCR. The primers used were 28F and 1492 R for the 16S rRNA ribosomal gene. Reagents used for the PCR reaction were DreamTaq PCR Master Mix (1X), dNTPs (2 mM), 5 pmol of each primer, and Taq polymerase (0.25 µl). The amplification was performed with an initial step of denaturation at 95°C for 5 min, then 35 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds and elongation at 72°C for 1.5 min. The final elongation step was at 72°C

for 5 min. The PCR products were purified using the Zymo ZR-96 DNA Clean-Up kit and the products were sequenced by Microsynth, Vienna. The sequences were identified using BLAST[48].

2.5. PCR clamping using Locked Nucleic Acids (LNAs)

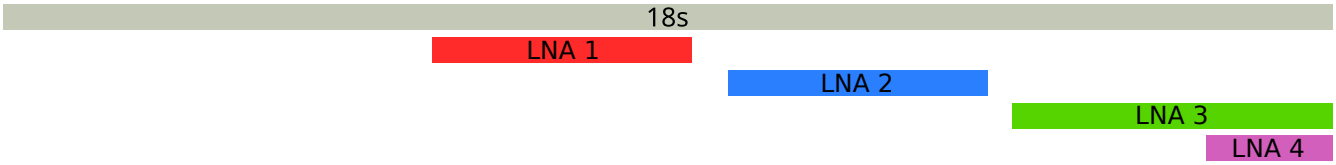


Figure 1: Schematic of the Locked Nucleic Acids (LNAs) used in this project. Shown at the top is the *Waminoa* 18S segment in gray. LNAs 1-4 are placed in parallel where they align to the segment.

To attempt to reduce interference from incidental amplification of host 18S DNA, I used new LNAs that we created specifically for this *Waminoa* system. I aligned the previously identified 18S sequence (see below) and the ten most prevalent identified 16S sequences from animal samples using MAFFT[49]. Four interesting regions that contained higher numbers of sequence divergence from the 16S were chosen (Figure 1) and ordered as oligomers with 3- or 6-carbon-chain spacers from Biomers. The sequences are as follows:

LNA 1:
5’-ATCGACCATAAGGGATAGTCTTTCAGATTATTCTGGTCACGGA -3’
end modification: 3'-spacerC3

LNA 2:
5’-TTTTACTTGTTAGACAATTTGCACTTAATTGTGTAGATTGATC-3’
end modification: 3'-spacerC3

LNA 3:
5’-CAAGTAACAGACCTTGAGAAAATTTGAGTGCTCAAAGCAAGCTTTGGCATTGA-3’
end modification: 3'-spacerC3

LNA 4:
5’-CAAAGCAAGCTTTGGCATTGA-3’
end modification: 3'-spacerC6

I initially tested the LNAs using gradient PCRs at different annealing temperatures using an animal tissue sample previously extracted with the DNeasy PowerSoil Pro kit as a template. Concentrations of reagents used in the PCR were as follows: 1X DreamTaq buffer (Thermo Scientific), dNTPs at 0.2 mM, primers at 0.25 µM each, DreamTaq (Thermo Scientific) at 0.625U, and PCR grade water

to a total of 25µl. Primers were 515F as per Parada *et al.*[50] and 806R as per Apprill *et al.*[51] PCR conditions were as follows: 95°C for 5 min; 35 cycles of 30s at 95°C, 60s at 50°C - 62°C, 90s at 72°C; then 72°C for 5 min, and storage at 12°C until removed from the cycler.

Due to the different sizes of the targeted 16S fragment and the interfering 18S fragment, I was able to verify results using gel electrophoresis with 1.2% agarose gels with Tris-Acetate-EDTA (TAE) chemistry run at 100 volts for 45 min. Based on these results, the temperatures chosen were: LNA 1 – 62°C; LNA 2 – 60°C; LNA 3 – 55°C. LNA 4 was not selected for sequencing due to its lack of effect in the tests.

For sequencing including the LNAs (see below), the following final PCR conditions were chosen for the first step PCR: 94°C for 3 min; 30 cycles of 45s at 94°C, 60s at LNA-dependent temperature, 90s at 72°C; then 72°C for 10 min, and stored at 4°C until removed from cycler.

2.6. 16S amplicon sequencing

Sequencing was performed by the Joint Microbiome Facility, Vienna. This facility uses a workflow published by Pjevac *et al.* in 2021[52]. Briefly, in this workflow, a two-step PCR is performed, the first step to amplify the target DNA, the second step to add a dual barcoding system. Next, samples are multiplexed and concentrated using the innuPREP PCRpure kit, and prepared and indexed using the TruSeq Nano DNA Library Prep kit and TruSeq Nano DNA Single Indexes Set A and B with alterations. Then, paired end sequencing is performed using an Illumina MiSeq. Finally, the barcoding system is used to demultiplex sequences, after which barcodes, linkers, and primers are trimmed, and sequence variants are called.

2.7. Bioinformatic analysis

For data analysis, I used R version 4.4.1[53]. I imported the data with the package DADA2[54] using scripts provided by the Joint Microbiome Facility. A large fraction of Forward and Reverse sequence reads could not be paired because they were incidental amplification of the host 18S fragment, which was too large to allow sufficient read overlap; therefore, I imported the raw reads as single reads for all results except the QIAamp DNA Microbiome kit results. Reads from that kit could be paired at a normal rate and were therefore imported as paired-end reads. In both cases, the reads were filtered for PhiX contamination with BBDuk[55]. The python package demultiplex[56] was used for demultiplexing (allowing 1 mismatch for barcodes and 2 mismatches for linkers) and primer removal (allowing 2 mismatches for forward and reverse primers each). BBDuk was again

used for trimming off barcodes, linkers and primers. Amplicon Sequence Variants (ASVs) found in extraction blanks and known facility contaminants were removed from the dataset.

Phylogeny was determined using SILVA datasets (version 138.1 for 16S data and version 132 for 18S data)[57] and DADA2, with scripts provided by the Joint Microbiome Facility. I used both the 16S dataset and the 18S dataset on every library, since neither is sufficient to reliably identify sequences as 16S or 18S on its own. ASVs identified at order level by only one dataset were counted as 16S or 18S respectively. I used BLAST[48] to further identify any ASVs identified at order level or higher with both datasets; any ASVs not sufficiently identified by either, I counted as 18S and 16S depending on their BLAST result with the highest E-value, and also confirmed that there were no conflicting results in the top ten result (which never occurred). From the resulting dataset, I calculated the percentage of 16S reads (excluding reads identified as mitochondria and chloroplasts) per sample, then removed 18S sequences, mitochondria sequences, and chloroplast sequences from the datasets for all further analysis. I also used BLAST to verify any ASVs of particular interest, such as dominant ASVs.

I used the R package vegan version 2.6-6.1[58] for further analysis. I generated rarefaction curves using the built in rarefaction function and visualized the output as described below.

For alpha diversity, I calculated the Shannon index of each sample, then removed all samples with zero alpha diversity for testing, since these were samples with not enough reads to find more than one ASV, and I expected them to heavily skew results. In principle, an arbitrary cut-off could have been set for how many reads a sample must have to be included, but since the alpha diversity was already partially compromised and reduced to only some of the samples, I decided to include everything with an alpha diversity above zero. After verifying normalcy using a Shapiro-Wilk normality test ($p = 0.17$ meaning null hypothesis of normal distribution is assumed) I determined whether differences between samples were significant using a two-way ANOVA based on method and worm line, and a TukeyHSD test to determine which factors impacted differences.

For beta diversity, PERMANOVA was calculated using adonis2, with Bray-Curtis distance as distance matrix. Validity of results was verified using the vegan package's betadisper function.

Finally, results were visualized using the packages ggplot2 version 3.5.1[59] and Polychrome version 1.5.1[60]. For clarity, labels were altered in large graphs to display the same information in a more readable way.

3. Results

3.1. Fraction of 16S reads in amplicon sequencing

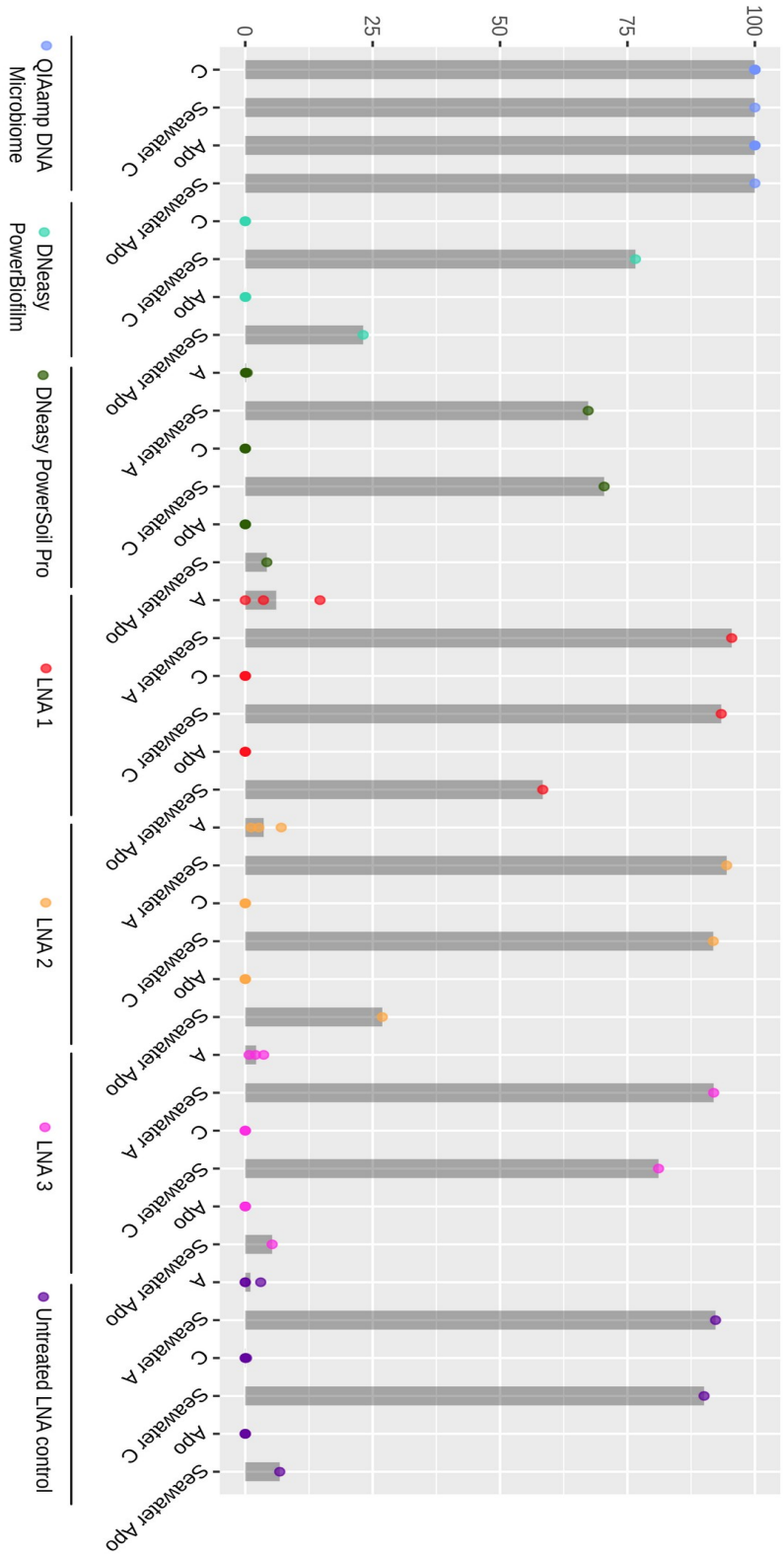


Figure 2: Percentage of reads identified as a 16S segment per category of sample (combination of method used and animal line/seawater control). Categories are sorted first on the x axis by method, then within method by animal line or seawater controls from different animal lines. Note that LNA treated samples were also extracted using the DNeasy PowerSoil Pro kit, but marked separately for easier visibility. Points (overlapping in many cases) represent unique samples, columns represent averages.

Animal tissue samples extracted with the DNeasy PowerSoil Pro and DNeasy PowerBiofilm kits produced amplicon sequencing results in which consistently over 99% of reads were incidental amplification of the host *Waminoa* 18S region (Figure 2). The total amount of reads from animal tissue samples ranged in the DNeasy PowerSoil Pro kit samples from 7262 to 24751 and in the DNeasy PowerBiofilm kit samples from 2893 to 19079. In these samples, the number of 16S reads was extremely low, ranging from 0 to 15 with one outlier of 51 reads, and the number of mitochondria- or chloroplast-identified reads ranged 0 to 17. The corresponding seawater samples extracted with these two kits produced much lower fractions of 2.47% to 40.49% 18S reads, with the number of 16S reads ranging from 609 (DNeasy PowerBiofilm) to 15222 (DNeasy PowerSoilPro), and total read numbers ranging 14379-25033 for the DNeasy PowerSoil Pro kit and 2632- 7396 for the DNeasy PowerBiofilm kit.

In contrast, samples extracted with the QIAamp DNA Microbiome kit (which has steps targeting host depletion) and sequenced produced no read pairs that could be identified as 18S (Figure 2). Different from the other results, paired-end reads were not converted to single reads in sequenced samples extracted from this kit, because the maximum loss of reads in the merging step in any animal sample was only 12.5%. There were also no reads identified as mitochondria or chloroplasts. The number of paired reads in animal tissue samples extracted from this kit was substantially higher than those extracted with the other two kits, ranging from 43-100 reads per sample. In a similar pattern, the number of paired reads in seawater control samples was also higher than in animal tissue samples, at 524 and 3327 respectively for the two samples examined.

The untreated (no-LNA) control for the LNA samples, which used the samples from the DNeasy PowerSoil Pro kit with no further treatment, produced a minimum of 96.97% reads identified as 18S. Out of 9 animal tissue samples, one sample produced one read and another sample produced 2 reads identified as 16S, and none of the other seven samples produced any read that was not identified as 18S, mitochondria or chloroplasts. The total number of reads ranged from 66-3807. Again the seawater samples resulted in more 16S reads: the fraction of 18S ranged from 1.29%-37.32% and the number of total reads ranged from 544-10510, of which 502-1102 reads were identified as 16S sequences.

The lowest fraction of 18S reads in animal tissue extraction samples treated with LNA 1 was 84%; this sample had 11 reads identified as 16S and was produced from an A-line worm. One 16S read and one mitochondrial read were in another A-line worm sample. For every other animal tissue sample, all of the produced reads were identified as 18S. The total number of reads in animal tissue

samples ranged from 6-78. The fraction of 18S reads in seawater samples ranged from 1.07-4.55% and total read numbers ranged from 22-4214.

In animal tissue extractions treated with LNA 2, up to 14 reads identified as 16S were produced from A-line worms. This makes the lowest fraction of 18S found in animal tissue samples in this condition, at 91.76%. Other than a singular read identified as mitochondria in a symbiotic C-line worm sample, all symbiotic and aposymbiotic C-line worm samples had only reads identified as 18S. Total read numbers ranged from 59-525. Seawater sample sequences ranged from 0.73% to 15.1% for fraction of 18S reads, and from 235-755 for total number of reads.

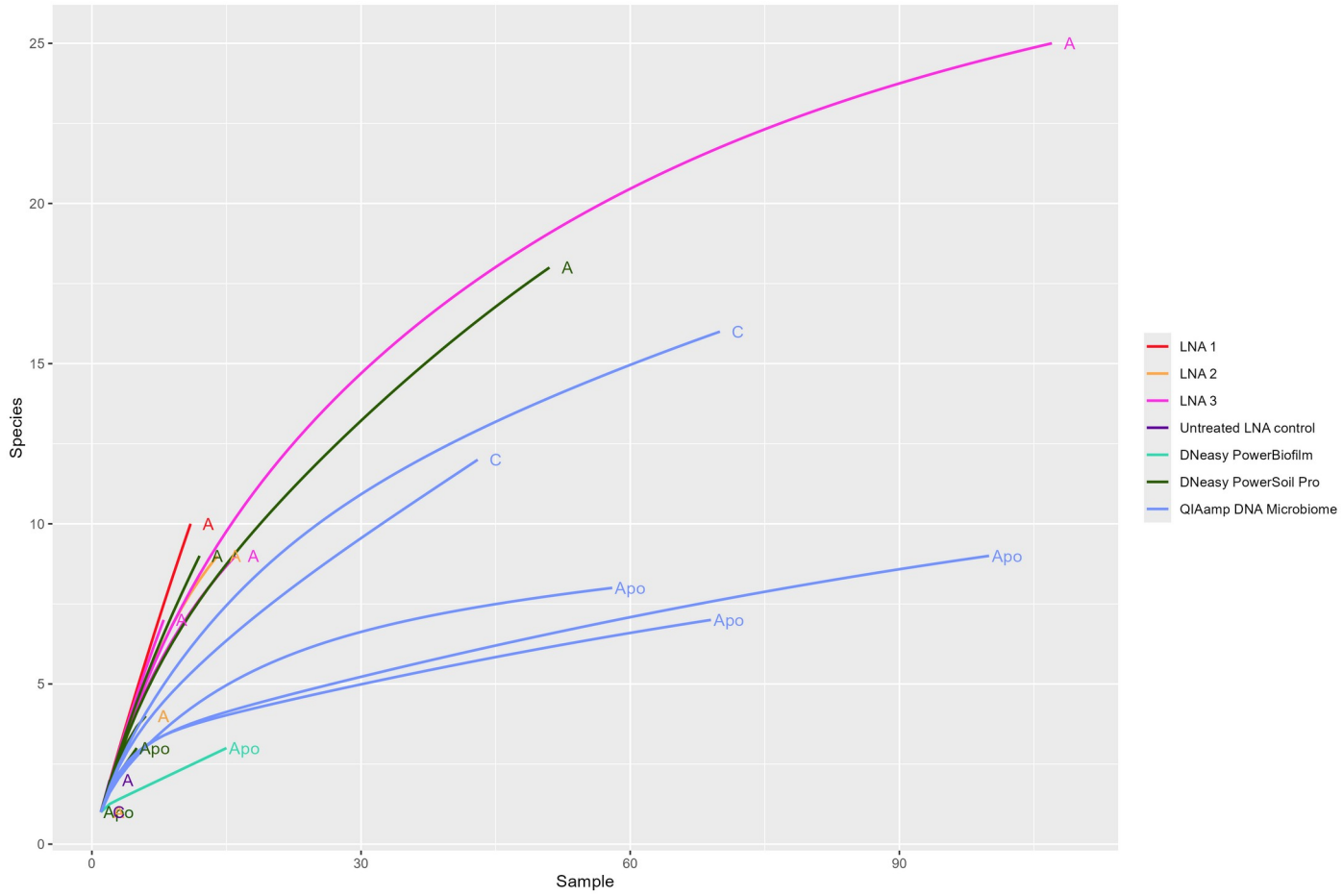
The fraction of 18S reads in A-line worm tissue samples treated with LNA 3 ranged from 96.53% to 99.2% (number of 16S reads were 8, 16 and 106 respectively, with no more than 1 read per sample identified as mitochondria). Other animal tissue extraction samples treated with the same LNA produced no reads identified as 16S (two reads were identified as mitochondria in an aposymbiotic C-line worm sample). The number of total reads produced from animal tissue samples ranged from 161-2959. Seawater samples ranged from 2.74% to 44.43% 18S, with total read numbers ranging from 2350-10593.

Taken together, animal tissue samples extracted with the DNeasy PowerBiofilm kit and the DNeasy Powersoil Pro kit, including extractions from the latter treated with any of the three LNAs, all produced extremely high proportions of 18S reads and correspondingly extremely low (or even no) 16S reads. Of those, A-line worm samples produced slightly more 16S reads than their aposymbiotic or symbiotic C-line worm sample counterparts. However, the seawater control samples extracted from these kits produced higher proportions of 16S reads. In contrast, samples extracted with the specialized QIAmp Microbiome kit produced only 16S reads, with no detectable 18S read contamination (Figure 2).

3.2. Sequencing depth

Overall, the rarefaction curves (Figure 3) of animal samples showed not enough sequencing depth to reach plateau. Samples produced from A-line worms show longer, but regardless steep curves, in comparison to their symbiotic and aposymbiotic C-line counterparts. Samples extracted using the QIAmp Microbiome kit are flatter, but do not reach plateau either.

As for the animal tissue samples treated with each of the LNAs, all indicate very low sequencing coverage. A singular sample treated with LNA 3 shows slightly deeper sequencing, but remains


Species
Sample
0
5
10
15
20
25
0
30
60
90
A
C
Apo
LNA 1
LNA 2
LNA 3
Untreated LNA control
DNeasy PowerBiofilm
DNeasy PowerSoil Pro
QIAamp DNA Microbiome

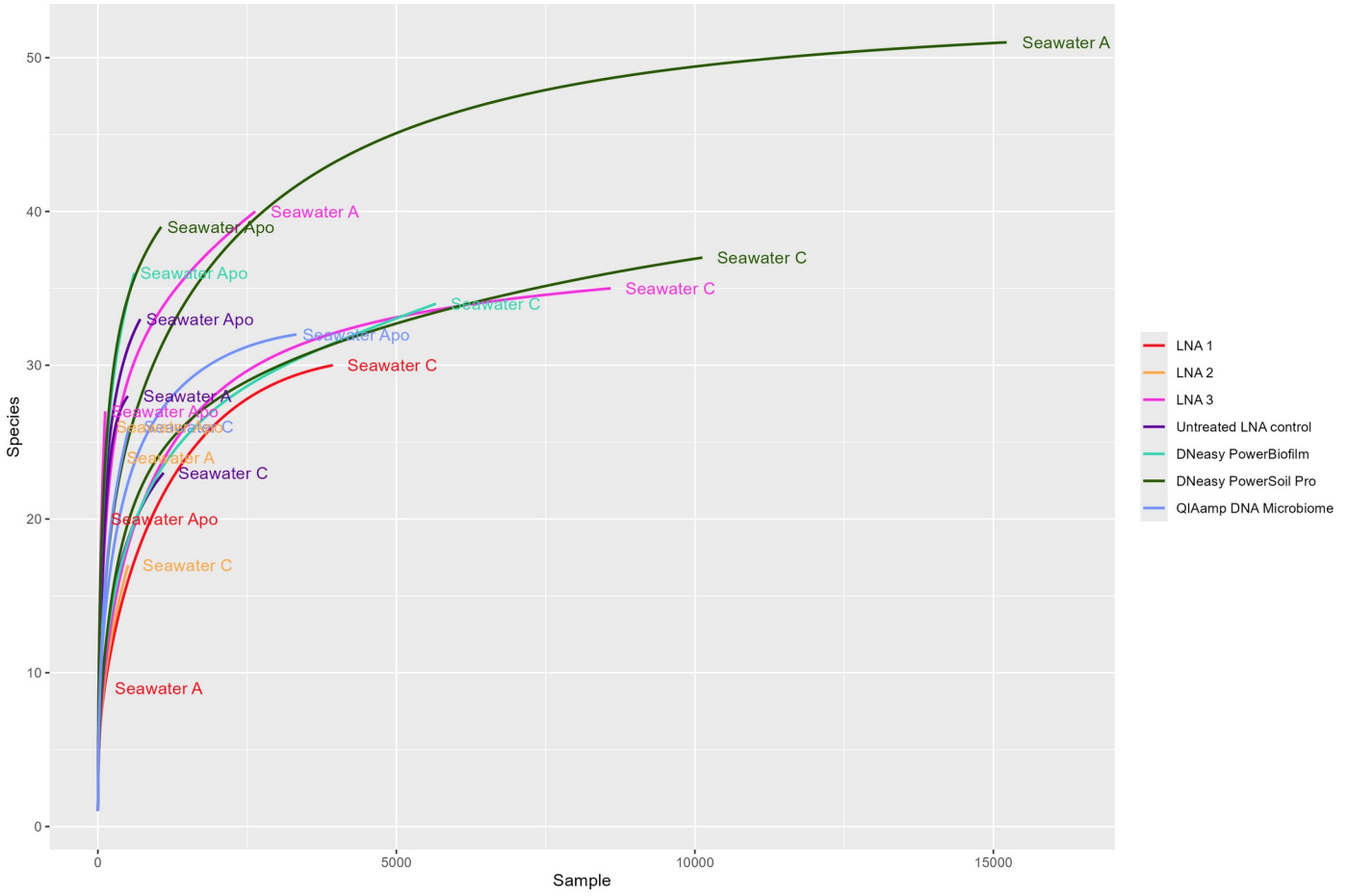

Species
Sample
0
10
20
30
40
50
0
5000
10000
15000
Seawater A
Seawater C
Seawater Apo
LNA 1
LNA 2
LNA 3
Untreated LNA control
DNeasy PowerBiofilm
DNeasy PowerSoil Pro
QIAamp DNA Microbiome

Figure 3: Rarefaction curves of all sequenced samples. Color denotes method, and additional labels indicate animal lines. Animal samples (*upper panel*) and seawater control samples (*lower panel*) have been placed on separate graphs for better visibility due to the large difference in scale (note the different x axis).

shallow. Comparing the untreated LNA control sample to its corresponding samples in the initial DNeasy PowerSoil Pro sample sequencing shows that the rarefaction curve is smaller.

As for the seawater control samples, rarefaction curves show orders of magnitude higher numbers than their animal tissue counterparts, but ultimately do not reach plateau in most cases, suggesting that deeper sequencing would reveal more diversity in these communities. DNeasy PowerSoil Pro samples show the strongest sequencing depth. Other seawater samples are inconsistent within the same kit and/or treatment, forming no discernable pattern.

3.3. Alpha Diversity of microbial communities

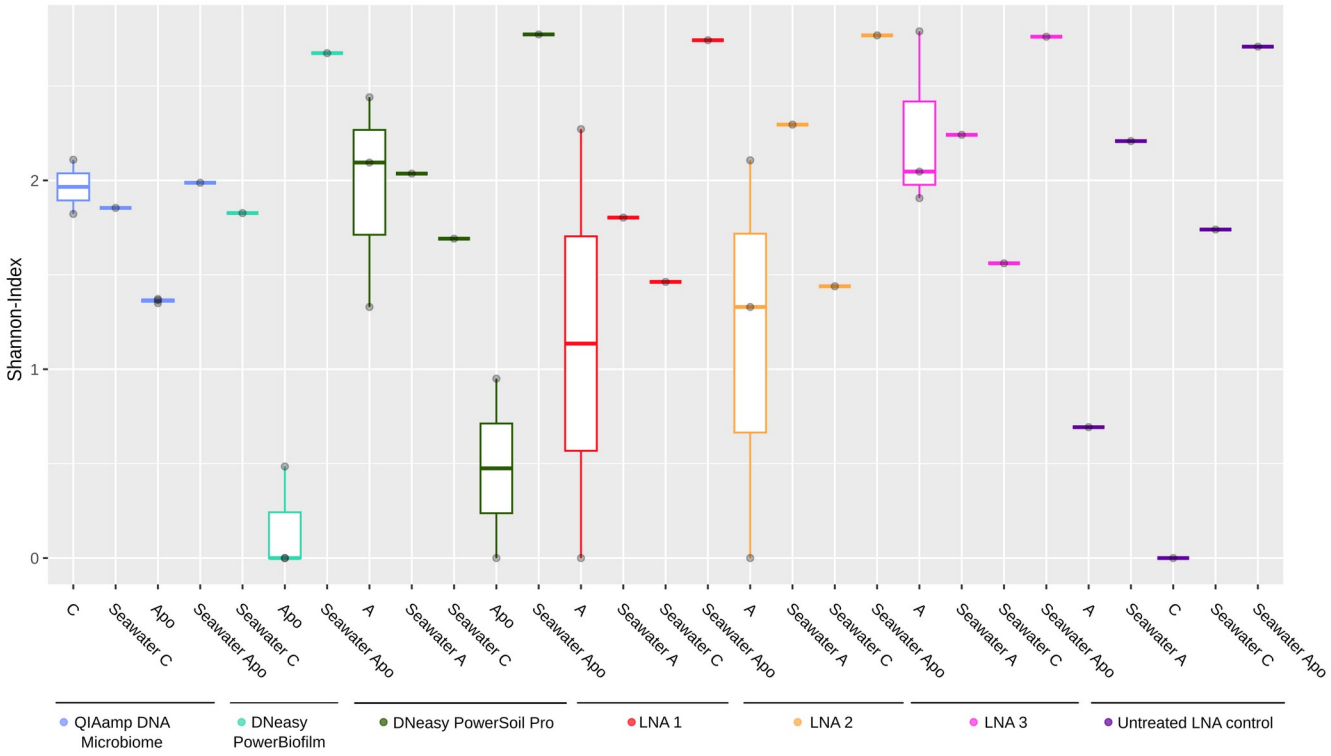


Figure 4: Alpha diversity of all samples, combined into categories of method and animal line. Note that samples in which no 16S signal was detected could not be included. Single lines rather than boxplots appear in categories that have only one non-zero sample (such as seawater controls) or when all replicate samples overlap.

Overall, alpha diversity shows little consistency across samples (Figure 4). There is a significant difference between animal lines across all viable (Shannon-Index > 0) samples (ANOVA, $p < 0.001$) and no significant difference between methods ($p = 0.37$). This significance, however,

stemmed from the seawater samples from aposymbiotic C-line worms, specifically the difference of the seawater to the aposymbiotic worm (TukeyHSD, adjusted $p < 0.001$), to the seawater of the symbiotic C-line worms (adjusted $p = 0.003$), and to the A-line worms themselves (adjusted $p = 0.006$), but not their corresponding seawater sample. Despite significance, these results are based on a very low number of seawater samples, and for the sake of certainty, this test should be repeated with a larger number of samples.

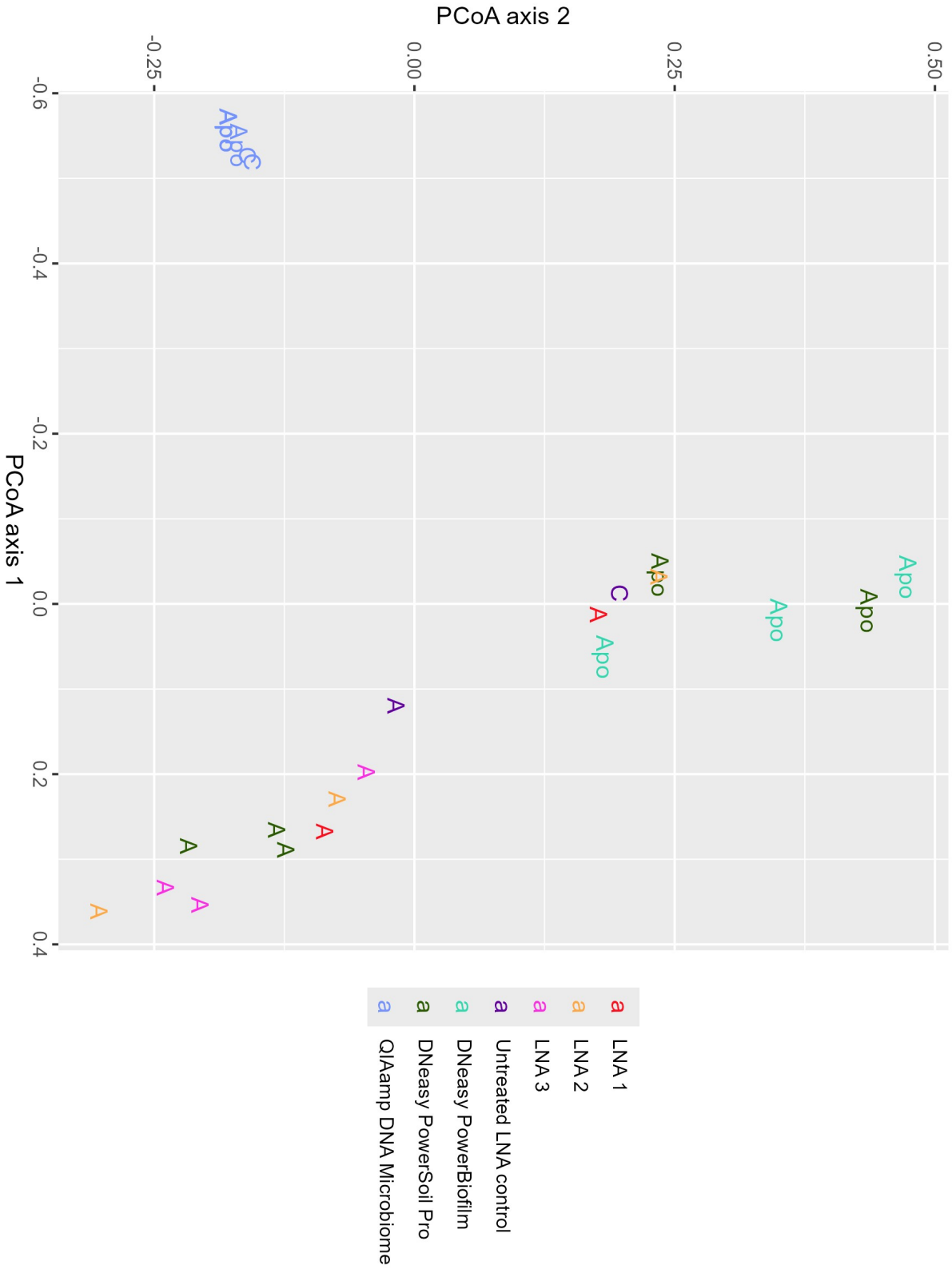
3.4. Beta Diversity of microbial communities

For animal tissue samples, a two-factor PERMANOVA test (animal line + method) revealed significant differences between different methodologies ($p = 0.001$, $R^2 = 0.4639$), but not between different animal lines ($p = 0.054$). However, when testing betadispersion using the vegan package's betadisper function, the test for different methods returned a significant ($p = 0.014$) value. For this reason, the validity of the PERMANOVA test may be limited. In the PCoA plot of beta diversity among the samples (Figure 5), samples from the QIAamp DNA Microbiome kit cluster together in particular, so as a control, I repeated the PERMANOVA without these samples. The betadispersion test returned not significant ($p = 0.548$) for methodologies. Both animal line and methodology were significant in the new two-factor PERMANOVA (animal line: $p = 0.014$, $R^2 = 0.1534$; method: $p = 0.025$, $R^2 = 0.3354$). The R^2 of the methods outranked the R^2 of the animal line, implying that even without the QIAamp DNA Microbiome kit samples, the choice of method had a greater influence on the results than the choice of animal line. However, with the QIAamp DNA Microbiome kit removed, the betadispersion test for the animals returned significant ($p = 0.001$). Therefore, the validity of these results may still be compromised.

The same tests were used to compare animal lines directly. A comparison between symbiotic C-line worms and A-line worms, which are always symbiotic, returned significant ($p = 0.002$, $R^2 = 0.0881$), with a non-significant beta dispersion ($p = 0.4$). On the other hand, comparison between symbiotic and aposymbiotic C-line worms using the data from the QIAamp DNA Microbiome kit was not significant ($p = 0.1$). There was not enough data (ie. too few 16S reads) among samples from C-line worms extracted from other methods to make further comparisons.

A PERMANOVA using only the LNA-treated animal tissue samples and the untreated LNA control revealed no significant difference ($p = 0.095$) between the different LNAs and the control (betadisper for methods: $p = 0.823$).

The seawater control samples are included in this section for comparison to animals; however, there are not enough samples for statistical testing. While there are possible differences, a greater sample size would be needed to determine if these differences are significant and replicable.



(a)

(b)

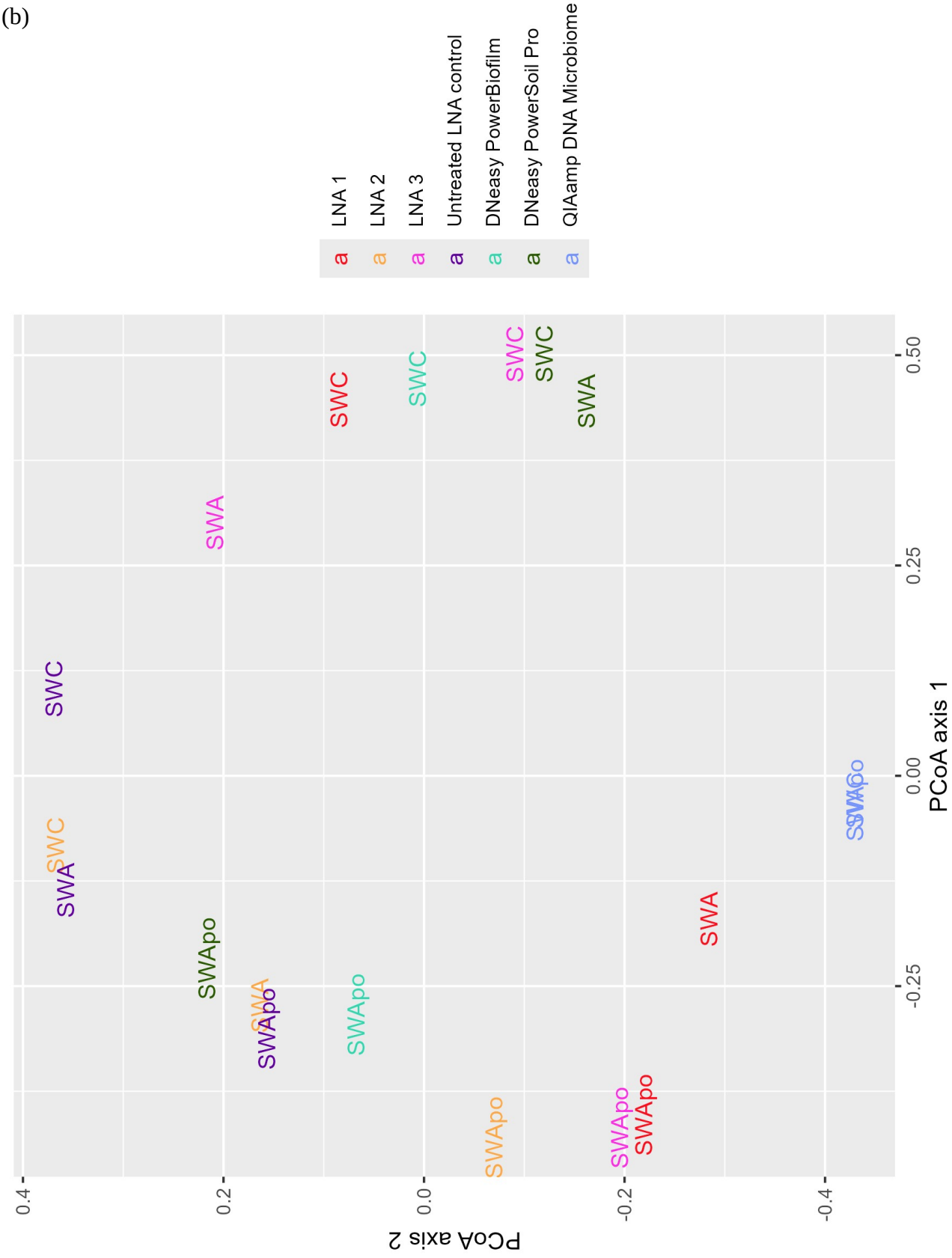


Figure 5: PCoA plots of animal samples (a) and seawater control samples (b), with sample origins as labels and method denoted by color. Seawater is abbreviated as “SW”. The distance metric used is Bray-Curtis.

3.5. Microbial community composition in *Waminoa* and seawater

3.5.1. ASV level

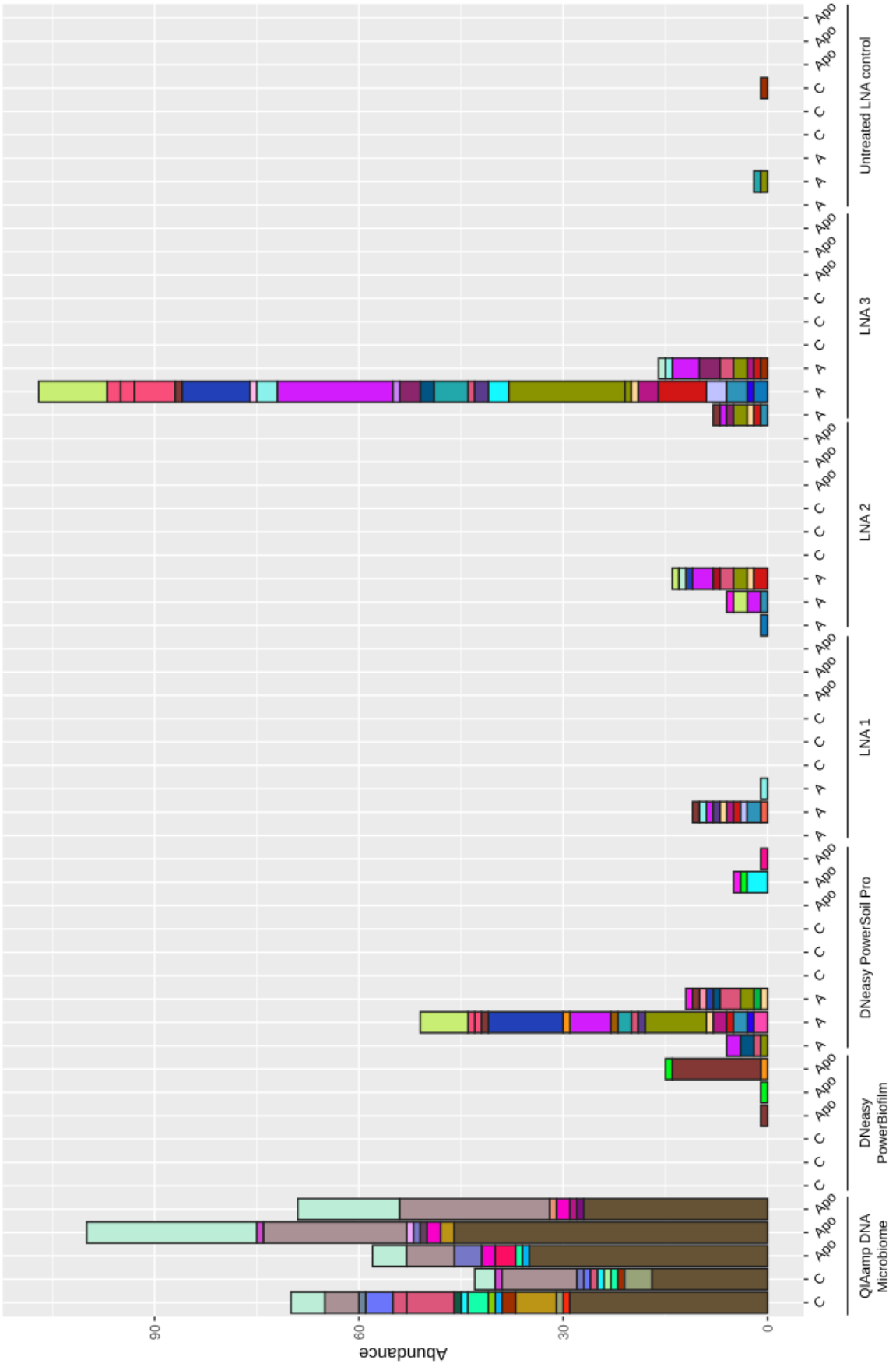
The compositions of communities in animal tissue samples at the finest resolution, the ASV level, do not agree between methods (Figure 6a). This is partially due to a lack of any reads identified as 16S in 33 samples out of the 56 total animal samples (Figure 2). In the remaining 23 samples, only 12 samples have more than ten 16S reads, meaning that a high stochasticity in the results from these very different total read amounts is also likely. In total, 65 unique ASVs were detected in animal samples.

DNeasy PowerSoil Pro and DNeasy PowerBiofilm samples only produced reads identified as 18S, chloroplasts or mitochondria from symbiotic C-line worms. ASVs produced from the same kits in aposymbiotic C-line worms show no consistency. In samples of A-line worms extracted using the DNeasy PowerSoil Pro kit, results are inconsistent among all three replicates.

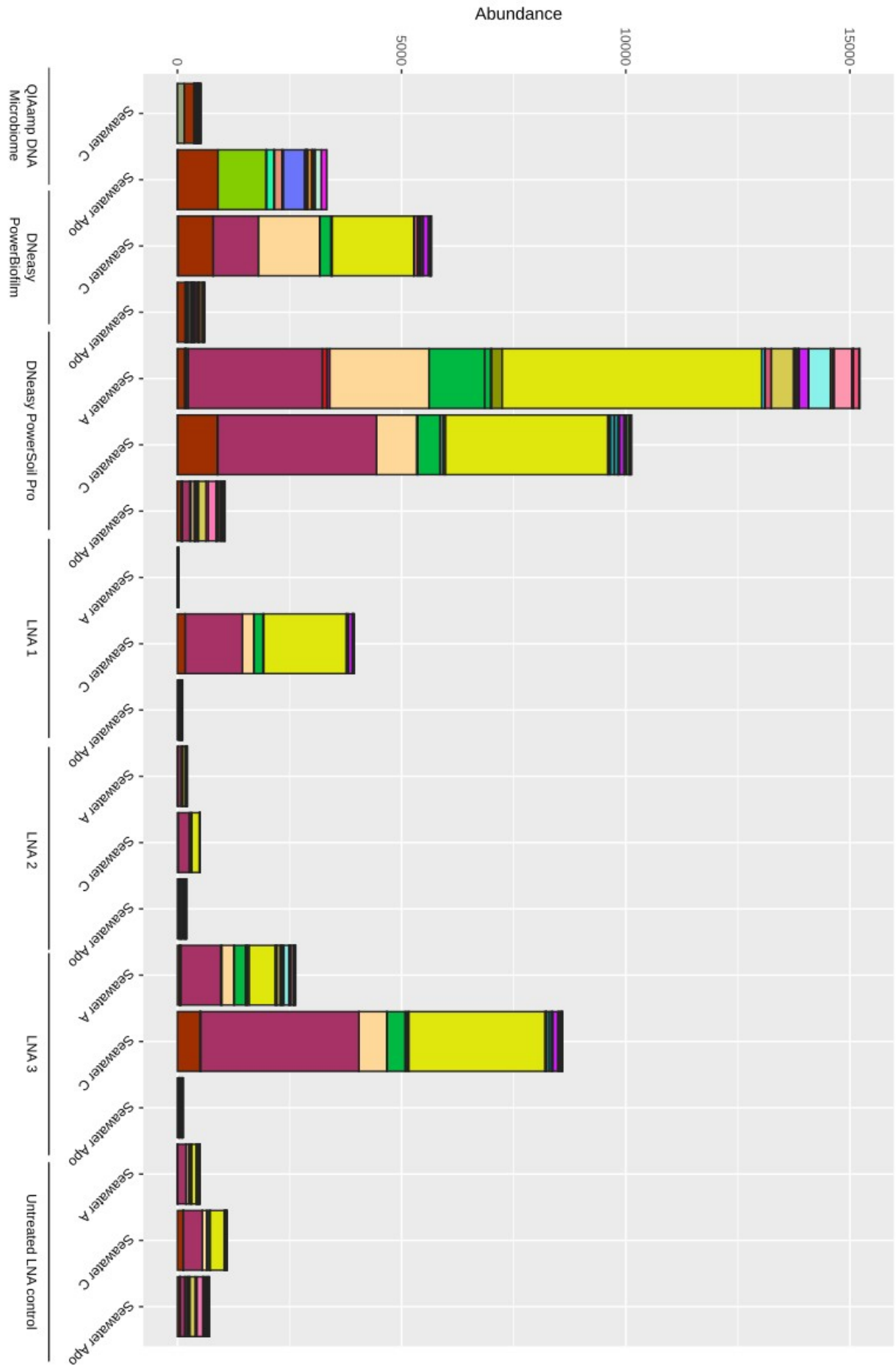
In contrast, within results produced from the QIAamp DNA Microbiome kit, the community composition is consistent, to a degree. Three ASVs are found dominant in all samples, though their fraction of the community differs. These ASVs were identified as *Alcanivorax borkumensis* (ASV_8mk_s14), *Yoonia-Loktanella sp* (ASV_f7s_c7z), and *Brevirhabdus pacifica* (ASV_tjl_k5n) by DADA2, although ASV_f7s_c7z was identified as *Sulfitobacter pontiacus* (NCBI Accession Number MK128666.1) when verifying via BLAST, and the identity of *Brevirhabdus pacifica* could not be verified due to a lack of results on BLAST. *Brevirhabdus pacifica* is noted to have a similar genome to the genus *Sulfitobacter*[61], which has been found and verified in the dataset, so this is a possible misidentification.

All three taxons are gram negative and aerobic[61–63]. The growth optimum of *Alcanivorax borkumensis* and *Sulfitobacter pontiacus* are listed as 25-30°C[62] and 30°C[64] respectively, fitting the 25°C the animals are kept at. *Alcanivorax borkumensis* was first described through screening for n-alkane-degrading bacteria in seawater and sediment in the North Sea[62]. It is primarily notable for its potential in bioremediation and association with oil spills, not for any sort of symbiosis[65]. *Sulfitobacter pontiacus* was initially isolated from H_2S-O_2 interfaces in the Black Sea and identified as a sulfite oxidizer[63], but it has been found in association with microalgae, which it protects from pathogenic bacteria[66].

(a)


Abundance
90
60
30
0
QIAamp DNA Microbiome
DNeasy PowerBiofilm
DNeasy PowerSoil Pro
LNA 1
LNA 2
LNA 3
Untreated LNA control
Apo
C
A

(b)


Abundance
0
5000
10000
15000
QIAamp DNA Microbiome
DNeasy PowerBiofilm
DNeasy PowerSoil Pro
LNA 1
LNA 2
LNA 3
Untreated LNA control
Seawater C
Seawater Apo
Seawater C
Seawater Apo
Seawater A
Seawater C
Seawater Apo
Seawater A
Seawater C
Seawater Apo
Seawater A
Seawater C
Seawater Apo
Seawater A
Seawater C
Seawater Apo
Seawater A
Seawater C
Seawater Apo

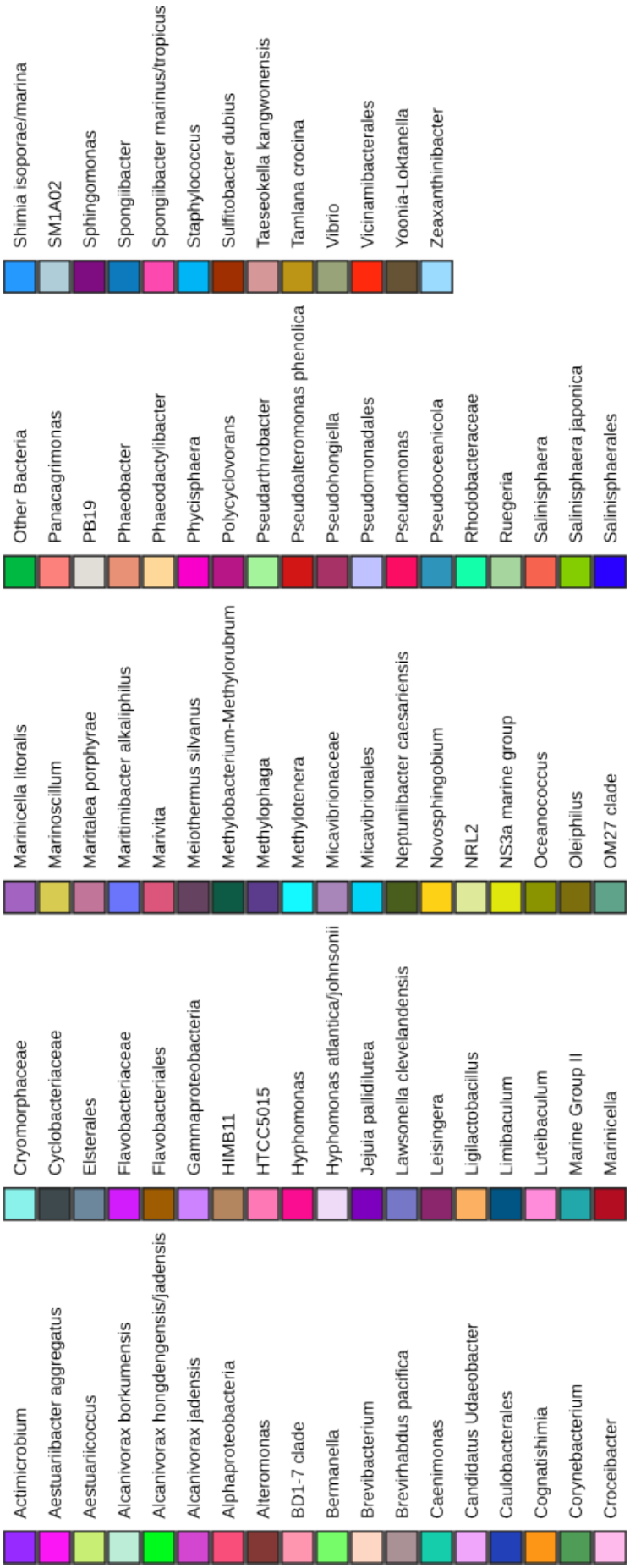


Figure 6: Absolute abundances of ASVs in all samples sequenced. Due to the large difference in scale, animal samples (*a*) and seawater control samples (*b*) are placed on separate graphs (note the different y axis), but shared ASVs are the same color in both graphs. ASVs identified as eukaryotic in origin have been removed.

LNA-treated samples produced no reads identified as 16S from any symbiotic or aposymbiotic C-line worms. LNA 1 produced prokaryotic ASVs only in two A-line worm samples, with no notable consistency between the two samples (one ASV with one read in each sample is the only overlap). LNA 2 produced communities for the three different A-line worm samples that are likewise inconsistent with each other. The compositions of the three A-line worm samples treated with LNA 3 are inconsistent with each other; however, interestingly, the most diverse sample of these has notable overlap with the most diverse A-line worm sample of all the DNeasy PowerSoil Pro samples (Figure 6a). The dominant overlapping ASVs were identified as *Caulobacterales sp* (ASV_7rs_jqt), *Flavobacteraceae sp* (ASV_d3o_e7x), *Aestuariicoccus sp* (ASV_pyb_tcq), and *Oceanococcus sp* (ASV_rcm_zh6) by DADA2, although these could not be verified with BLAST due to a lack of similar sequences. In the order in which they have been named, the read counts for each are, in the LNA 3 treated sample, 10, 17, and 10, respectively; in the DNeasy PowerSoil Pro samples, the numbers are 11, 6, and 7 respectively. As for the no-LNA control, there are only three reads in total, identified as a different ASV each.

Members of the three identified genera were isolated from marine environments[67–69], and the World Register of Marine Species describes the order of Caulobacterales as being marine[70]. Only the genus *Flavobacteraceae* has been described as commonly living in close association with animals, through which it takes a variety of roles[68].

Across all methods, symbiotic C-line animal samples rarely produced ASVs identified as prokaryotic 16S. Outside of a singular ASV produced in the no-LNA control, only the QIAamp Microbiome kit extracted samples resulted in 16S ASVs. Aposymbiotic C-line animals, on the other hand, produced a small number of ASVs in samples extracted with the DNeasy PowerSoil Pro kit and the DNeasy PowerBiofilm kit, though none in any sample treated with LNA, or its no-LNA control. A-line worms were only extracted using the DNeasy PowerSoil Pro kit, and the resulting samples were sequenced both without further treatment and with LNA treatment. ASVs identified as prokaryotic 16S were found in 12 out of 15 samples, often in small numbers and different composition. The most consistent ASVs were identified as described above.

Seawater samples show some reappearing taxa, with the caveat of the sample numbers being low. Overall, the most abundant taxa in seawater are *Phaeodactylibacter sp* (ASV_ecm_s4s), *Pseudohongiella sp* (ASV_pq3_zrd), and *NS3a marine group sp* (ASV_tnd_joh) (Figure 6b). These identities could not be verified with BLAST, as none of them had any matches with significant similarity. Seawater control samples from aposymbiotic animals appear to have a different

community, although the number of samples is not sufficient to test for statisticially significant differences.

Between non-empty animal samples and seawater samples, there is notable overlap. ASVs found in symbiotic C-line animals were up to 30.2% (15 ASVs total) unique in samples extracted using the QIAamp DNA Microbiome kit, in contrast, aposymbiotic C-line animal samples extracted using the same kit are at most 6% unique to the animals. Two aposymbiotic C-line animal samples extracted using the DNeasy PowerBiofilm had one ASV not found in seawater each, no other ASVs unique to animals was found in any other C-line animal samples (note that most other samples had very low numbers of ASVs identified as prokaryotic 16S, or were devoid thereof). As for A-line animals, the maximum number of ASVs found in any animal sample that was not found in any seawater sample was 10 (sample treated with LNA 1) and in 5 samples, all 16S ASVs detected were also detected in seawater samples.

3.5.2. Order level

For further analysis, I have hereafter excluded samples with no reads identified as prokaryotic 16S. I also will show community results using relative abundance; please note, however, that the relative abundances can be misleading in samples that have few reads.

On the order level, animal tissue samples remain inconsistent among replicates, but seawater samples begin to be distinguished between differently produced samples (Figure 7). Common taxa across all samples are Rhodobacterales, Flavobacteriales and Pseudomonales. Rhodobacterales occupy particularly large proportions of all samples extracted using the QIAamp DNA Microbiome kit. While Salinisphaerales and Rhodobacterales routinely show higher proportions in animal tissue samples (other than samples using the QIAamp DNA Microbiome kit), and Pseudomonales tend to occupy larger portions of seawater samples compared to animal tissue samples, this may be due to the stark difference of total read numbers between animal tissue samples and seawater samples. Cytophagales and Phycisphaerales are orders that show up primarily in seawater samples from aposymbiotic C-line worms, as far as this set of samples is concerned.

3.5.3. Class level

Across all samples, the most common classes were Alphaproteobacteria, Gammaproteobacteria and Bacteroidia (Figure 8). In an interesting pattern, Phycisphaerae were only found in samples from seawater from aposymbiotic worms as well as in aposymbiotic worms extracted using the QIAamp DNA Microbiome kit. Actinobacteria are found almost exclusively in samples extracted using the QIAamp DNA Microbiome kit.

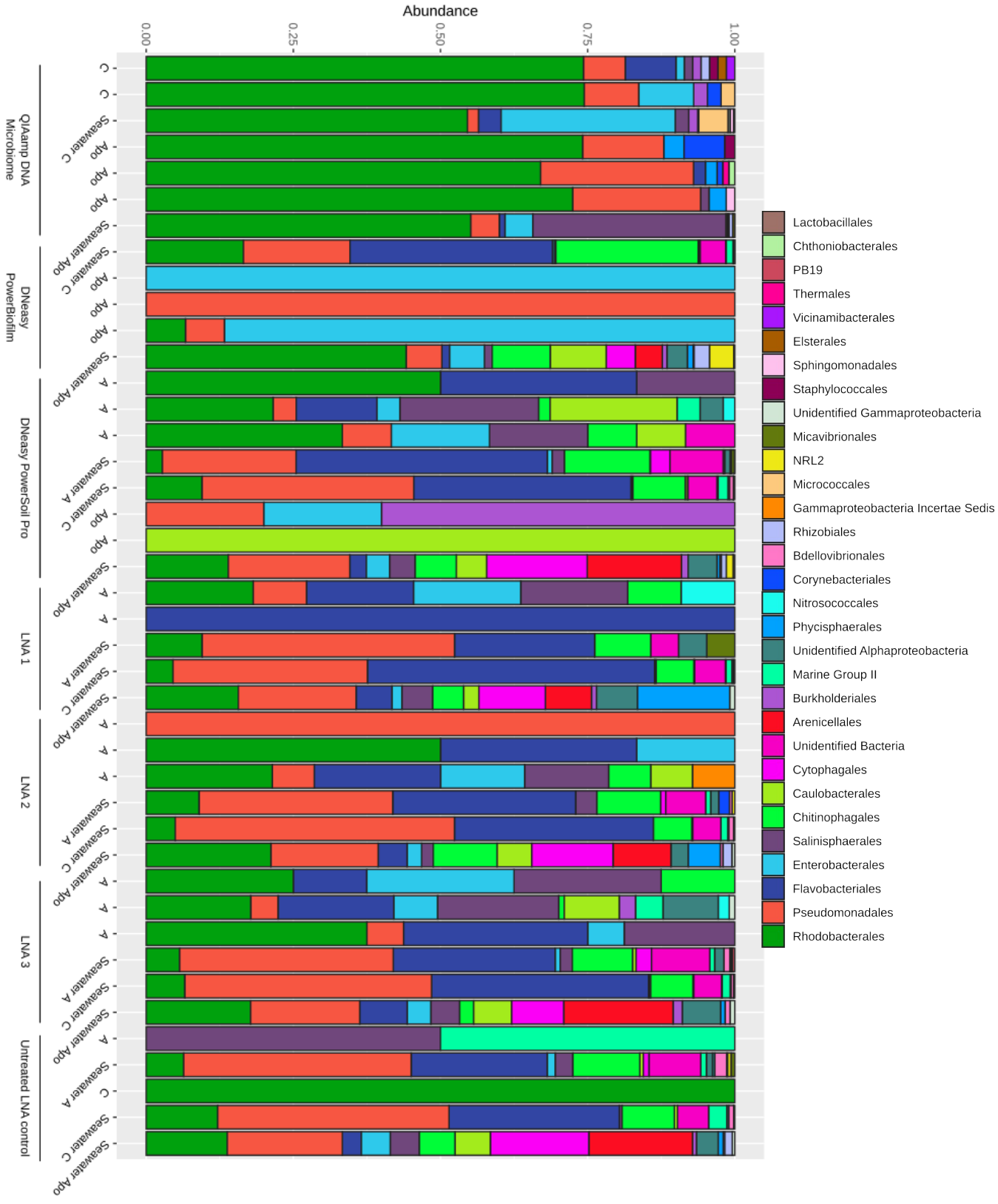


Figure 7: Relative abundance of prokaryotic orders. Note that samples in which no 16S signal was detected have been removed.

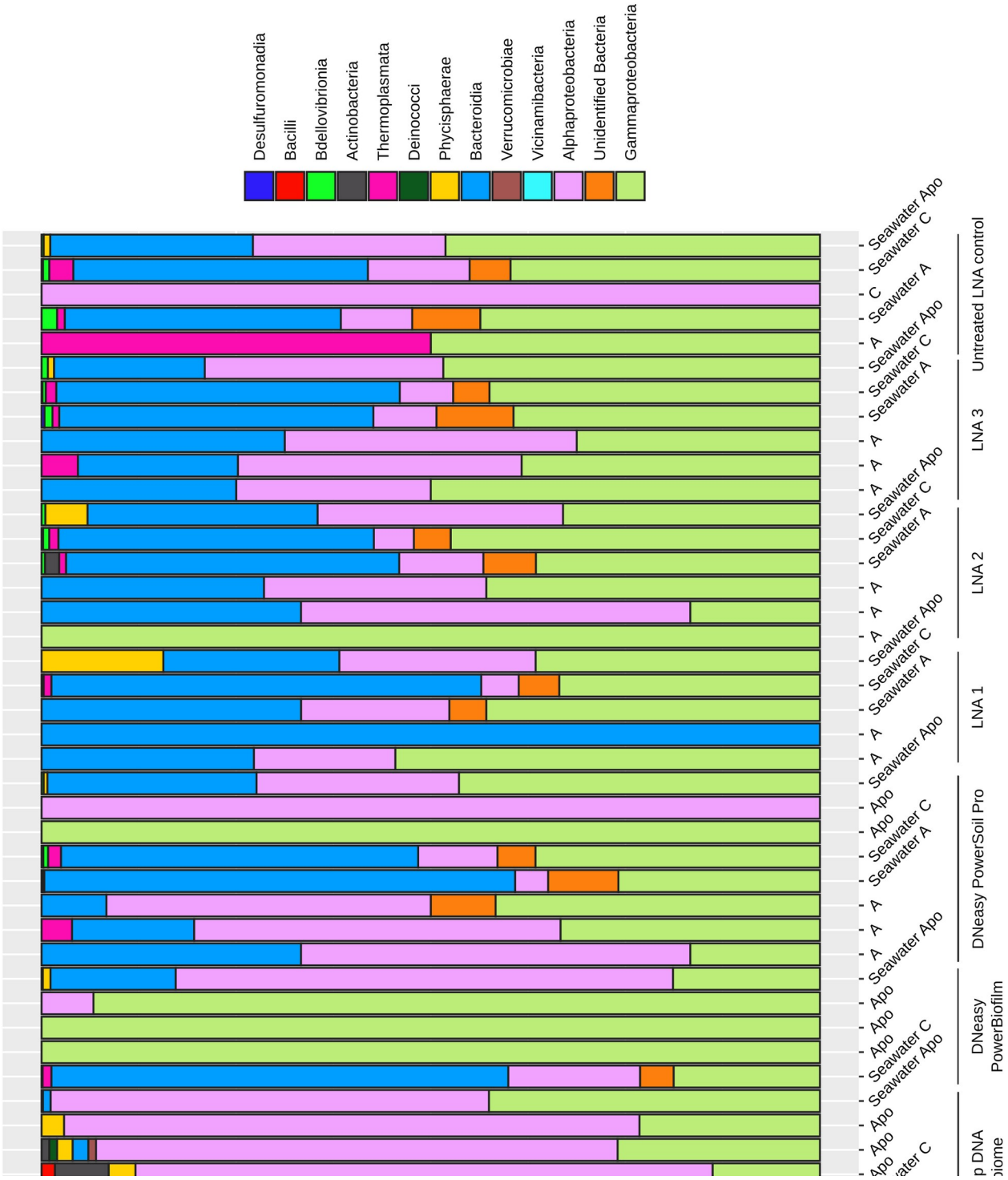


Figure 8: Relative abundance of prokaryotic classes. Note that samples in which no 16S signal was detected have been removed.

3.6. DNA extraction methods removed after initial testing

Not all tested extraction methods were sequenced. Certain methods were ruled out after initial testing (see Material and Methods). Physical methods of microbial material separation, such as centrifugation and filtering, resulted in low amounts of DNA after extraction (less than 5 ng/µl, and frequently less than 2 ng/µl). Gels of the products of PCR amplification of these templates also showed bands primarily at the length of the *Waminoa* 18S segment, which is longer than the bacterial 16S segment amplified.

As outlined above in the Materials and Methods, additional DNA extraction kits that I tested were the ZymoBIOMICS DNA Miniprep kit, the Qiagen DNeasy PowerSoil Pro kit with an altered protocol that extended the chemical lysis step to 24h, and the Zymo HostZERO Microbial DNA kit. The ZymoBIOMICS DNA Miniprep kit and the altered Qiagen DNeasy PowerSoil Pro kit showed no differences to the unaltered Qiagen DNeasy PowerSoil Pro kit in gels of their PCR products, and were therefore no further used. The Zymo HostZERO Microbial DNA kit extracts produced Qubit fluorometer measurements that were too low for detection, suggesting failure of DNA extraction, and were therefore also discarded. For more information, see Appendix Table 1.

3.7. Comparison to isolated bacterial strains

The analysis of the bacterial isolates was performed in cooperation with Sandra Montaño Salazar, a PhD student in our group. Using BLAST, three of the four strains isolated were identified as *Alteromonas sp* (NCBI Accession Numbers AB571943.1, HQ161365.1, AB571943.1) and the other was identified as *Sulfitobacter pontiacus* (NCBI Accession Number MT510151.1). The seawater control isolate was identified as *Vibrio sp.* (NCBI Accession Number OQ553742.1). The same strain of *Alteromonas sp* was found in seawater controls for every animal line (though not every single control sample) and in aposymbiotic C-line animal samples (DNeasy PowerBiofilm kit) as well as A-line animal samples (DNeasy PowerSoil Pro kit, LNA 1 treated, LNA 3 treated) (Figure 6a). It has been described as a free-living marine bacteria that occupies a variety of niches[71]. *Vibrio sp.* was found in symbiotic C-line animals (extracted using the QIAamp DNA Microbiome kit) as well as seawater controls for symbiotic and aposymbiotic C-line animals (QIAamp DNA Microbiome kit, DNeasy PowerSoil Pro kit, Dneasy PowerBiofilm kit) (Figure 6a). The DNA from *Vibrio sp.* also aligned closely with *Vibrio alginolyticus* (NCBI Accession Number CP046814.1), which has been found living in association with a marine sponge[72] and notably has been identified as a species producing toxins for the chemical defense of the pufferfish[73]. It is a known human pathogen[74]. The isolated strain of *Sulfitobacter pontiacus* aligned fully with ASV_f7s_c7z, which

had been identified as *Yoonia-Loktanella sp* by DADA2, and as *Sulfitobacter pontiacus* using BLAST (see section 3.5.1.) and was found in all animal samples extracted using the QIAamp DNA Microbiome kit.

4. Discussion

4.1. Effectiveness of host depletion in amplicon sequencing

The microbial amplicon sequencing of the animal tissue samples in these analyses was overwhelmed by incidental amplification of the *Waminoa* animal host's 18S DNA. While amplicons from mitochondria, chloroplast, and protist 18S sequences were present, it was only in small fractions and had no great impact on the resulting sequences. On the other hand, host 18S routinely made up over 90% in animal tissue samples, and frequently 99.9% in samples with no host depletion treatment. The resulting amount of actual microbial 16S sequences thus did not often reach the appropriate amounts needed for deeper analysis.

While eukaryotic host DNA is a frequent problem in the analysis of the host's prokaryotic microbiome, an imbalance of this degree is not common in the literature even in control groups. Ahannach *et al.* have examined host depletion methods in human samples and describe samples without targeted host depletion treatment to reach up to 95.9% human genome aligned reads[39]. Another study on host depletion methods used a control that comprised 6.7% bacterial DNA[32]. In a University of Vienna master's thesis on host organelle depletion in seagrass, about a third of the resulting reads were target DNA[75]. While these are low amounts, there is a stark difference between these reports of 4-7% target DNA, and the proportions 0.01% or less target DNA found in many of my samples. This could affect how effective host depletion strategies are.

Furthermore, the host depletion methods tested in this thesis have primarily been tested on human samples. This makes sense, as human host DNA can be a problem in human microbiome samples as well, and the recognition of the importance of the human microbiome in human health has drastically increased[76]. It is possible that these methods in their current state have, intentionally or not, been optimized for the human medical application. In his master's thesis on a symbiotic seagrass system, Eckensperger found similar problems with premade PCR clamps made from PNA when testing them on the seagrass host, as well as a lack of chloroplast depletion in kits specialized for depleting host DNA, and also noted the rarity of host depletion studies on plants[75]. Similarly, this is the first study of host depletion methods in the marine acoel worm *Waminoa*, and acoel worms in general have not been a primary target of host depletion applications. Available methods may therefore not be optimized and could require alteration for better effectiveness.

The reason why bacterial DNA in the *Waminoa* animal samples was so rare is not clear. As far as the animals are concerned, their microbiome has not been previously described. In nature, they do

not seem to receive algae symbionts from the corals they colonize[3]; whether the same is true for possible bacterial symbionts remains unclear. As kept in our facility, the animals have contact to little that could transfer bacteria into their system, because they are kept '*in vitro*' in glass bowls in filtered artificial seawater. This might have affected the community. Bathia *et al.* found that the freshwater medium used for their *Hydra* animals still affected the microbiome despite being filtered, and concluded that the effect likely was caused by phages and possibly metabolites that remained in the water after filtering[22]. The medium in which the *Waminoa* animals used in this thesis are kept is artificial seawater made with RO water and filtered to ensure that there are no bacteria, so a gap in this process that could introduce metabolites or phages altering the animal microbiome is not likely, but it is possible.

The low amount of bacterial 16S reads from *Waminoa* samples could also be related to the PCR process during amplicon generation. The outcome of a PCR is influenced heavily by which regions are targeted[25] and which primers are used[77]. Primers especially have complex interactions with DNA template that depend not only on the sequence of the primer itself, but on the PCR conditions and polymerase used as well[78]. It is possible that the primer used for PCR amplification for amplicon generation was more compatible with the *Waminoa* 18S sequence than other eukaryotic 18S sequences. Targeting a different region of the bacterial 16S gene is possible, but success is hard to predict. A PCR reaction containing multiple homologous templates is also highly competitive, and may preferentially amplify sequences depending on a variety of factors like primer binding capacity and GC content[77]. Even small alterations in PCR conditions can affect the amplification efficiency of a template[78]. Even when host depletion reduces host DNA, if the binding affinity is strong, it will still be likely a dominant part of the final result. While deeper sequencing of the samples here could therefore yield additional results, these results would likely still be overwhelmed by *Waminoa* 18S, and it would be an inefficient way to gain a small amount of new information. Bioinformatic analysis too depends on software and settings used[77]; however, within this study, the same analysis was applied to every sample and as the identification of the *Waminoa* 18S was a 100% match, there is little else that could have been learned from different software programs or settings.

In comparison to the animal samples, the seawater samples were affected differently by different DNA extraction methods. This is to be expected, as the three kits used for sequencing here are all designed for DNA extraction of animal tissues or larger amounts of organic material, not low amounts of material on filters. However, I also found *Waminoa* 18S DNA in the seawater, which could be caused by cells and particles that were broken off[79]. In all samples but those extracted by the QIAamp DNA Microbiome kit, the seawater control samples were very different from the

corresponding animal samples both in terms of proportion of identified 18S reads, and in terms of total amount of reads (Figures 2, 6). Samples extracted using the QIAamp DNA Microbiome kit did not include any reads identified as an 18S sequence, but the amount of reads in the seawater control exceeded the amount of reads in each animal sample in all cases. It is unclear whether this is due to a higher amount of bacteria in the water as compared to the animals, a lower amount of 18S DNA, or another effect in the animals that is not present in the water.

These findings align with other comparative studies: while not all kits have equal results, the QIAamp DNA Microbiome kit from Qiagen and the HostZERO Microbial DNA kit from Zymo Research show consistent results in different sample types[31,32,39,75]. While this is not true for other kits tested in these studies[31,32], none of those methods were tested within this thesis. Of note: When reporting the proportion of host and bacterial reads in terms of relative abundances, without reporting absolute read numbers as total abundances, essential information may be missing. While the fraction of bacterial 16S in samples in this thesis are generally speaking one-sided in most cases, when a study reports a higher fraction of bacterial 16S after host depletion, it is not clear if this indicates a higher amount of bacterial 16S reads or a lower amount of host reads with no changes in 16S reads. Indeed, one study that does report absolute abundances actually reports significantly lower amounts of bacterial DNA after host depletion in human skin samples[39]. Within this thesis, I also observed lower amounts of host DNA in some samples; for example, samples treated with LNA 1 showed a drastic reduction in the amount of host reads without an increase in the amount of reads from the bacterial community. Methods of host depletion are often more time- and cost-intensive than their non-specialized cousins, so more thorough reporting is absolutely necessary for good decision-making in the future.

4.2. Attempted blocking of host 18S amplification with LNAs

Despite the difference in DNA extraction methods being the greater influence on amplicon sequencing results according to PERMANOVA, different LNA treatments had no significant differences from each other or the untreated LNA control. This is despite the functionality of LNAs having been demonstrated as previously effective in blocking amplification of the LNA target[80], and has multiple possible reasons.

LNA design and optimization is a multi-step process. First, the LNA has to be designed. LNA blockers can take many forms: they can block either primer binding or elongation, can be purely LNA or a mixture of LNA and DNA, and can be of different lengths, all of which will affect their specificity and effectiveness[42] as well as melting temperature[46]. In addition to blocking, there is a

wide variety of undesired interactions of the LNA molecule with itself or other molecules[42] which need to be avoided through good design.

After LNAs have been synthesized, ideal PCR conditions for LNA usage must be found. The two factors I tested in this thesis were annealing temperature and concentration (see Materials & Methods), which makes for two continuous variables that require optimizing. Without spending time and money on sequencing, testing PCR products using gel electrophoresis is standard procedure, but the practical reality of gels can make determining size differences difficult. Furthermore, gels are only a viable option when the target DNA amplicon has a different size than the DNA amplicon that is targeted to be blocked. For testing many conditions via PCR, a large amount of DNA template is required, therefore if material is difficult or expensive to obtain, the process may not be feasible.

Finally, many alterations in preparing the sample and the PCR can also affect the functionality of the LNAs. This can include different DNA extraction methods, or anything that might affect the way primers or DNA behave during the PCR. Changes in primers might also affect primer-LNA interactions.

Another complication was introduced here: the template used for LNA-treated PCRs and subsequent sequencing. This was the same template as the DNeasy PowerSoil Pro kit samples that, after sequencing for the first time, were stored at -20°C for approximately 6 months. The untreated LNA controls were therefore the exact same DNA extraction samples, only after storage; despite this, the sequencing results were different. This could be a problem of storage, as long-term storage is generally recommended at -80°C[77], and different results after storage at -20°C has been demonstrated[81]. Regardless, it is unlikely that a new extraction would have yielded results in which the effect of LNAs would have been so fundamentally different that conclusions drawn from this thesis would have changed.

The process from initial LNA design to final effective LNA has many factors and comes with many decisions that need to be corrected and optimized. Within this thesis, it was not possible to complete this process by iterative design and testing via sequencing. This resulted in LNAs that did not have the desired effect in blocking amplification of host *Waminoa* 18S DNA. Interestingly, they did have an effect in that read numbers were affected by the presence of specific LNAs, but regardless of how much the amount of 18S reads was reduced, the amount of 16S reads was not increased. This could be due to the LNAs having off-target effects and reducing the desired DNA amplification alongside with the undesired DNA target. There is one sample that has a much greater amount of

reads that were identified as 16S compared to the rest (using LNA 3; Figure 2), but it is impossible to determine if this happened due to the presence of the LNA or stochastic factors in the PCR. Future studies could use more replicates to test whether this phenomenon occurs with LNA 3 treatment again, as well as continue to improve LNA design and effectiveness.

4.3. Alpha Diversity

While alpha diversity is considered a standard measurement of microbiome community diversity in samples, its interpretation is made difficult by the low read numbers and low sampling depth, as seen in the rarefaction curves and in the high variety in results above. For a start, while samples containing no reads identified as prokaryotic 16S were removed, any data point at 0 Shannon diversity means that only one ASV was found there, which as shown above is generally due to the low amount of reads (or even just a single read) rather than there being a true low diversity. Data points above 0 have at least two different ASVs, but are calculated from samples that have vastly different amounts of read numbers, with animal tissue samples routinely resulting in less than 100 reads and seawater samples in 1000 or more reads. Different sampling sizes and sampling depths can lead to incorrect conclusions about the true richness in an environment and cannot easily be corrected for[82]. Trying to correct using rarefaction is not an applicable solution in this case: samples with extremely low read numbers would have to be removed to create any sensible data at all (if all samples were to be included, this would mean rarefying down to one read per sample, placing every sample at a 0) and the cutoff would be arbitrary. This way, data would be shaped more by which samples are included rather than the data itself. Finally, the low amount of reads in most samples is not sufficient to represent the evenness of a sample, as the lack of repeated samples that is to be expected in larger numbers of reads might be misleading and result in an overestimate of diversity. This problem is reflected in the rarefaction curves, which for many samples end at a steep point, meaning that little of the found ASVs were repeated.

4.4. Community composition of the *Waminoa* prokaryotic microbiome

Throughout this thesis, samples extracted using the QIAamp DNA Microbiome kit have repeatedly shown the most striking differences from other samples. However, the low amount of reads present in many samples and the poor sequencing depth make it difficult to interpret resulting data. In samples with less than ten reads, results are too stochastic to make a definitive statement about the host animal’s bacterial community.

Outside of the QIAamp DNA Microbiome kit samples, there are two samples that have a sufficient amount of reads to make comparisons, though not definitive statements about the animal's bacterial community, due to their not being replicated within this study. Two samples from A-line worms, one extracted using the DNeasy PowerSoil Pro kit with no additional treatment and one extracted using the same kit with additional treatment with LNA 3, have similar dominant ASVs. This might be due to stochastic effects in PCR[30], such as primers by chance combining with bacterial 16S template early on and aiding its amplification, or a lower density of 18S DNA. Notably, these samples have little in common with the samples from the QIAamp DNA Microbiome kit. Unfortunately, this comparison is made difficult by the fact that there are no samples from A-line worms extracted using this kit. It is possible that the difference is due to different animal lines, but it is more likely that this is a result of using different methods, as differences in sequenced communities due to different DNA extraction methods has commonly been observed[37,38], including host depletion methods[33,75].

Naturally, this raises the question of which result is closer to the true microbial community, but this cannot be determined with the present data alone. Another comparison has found differences between the QIAamp DNA Microbiome kit and other host depletion methods in terms of resulting community composition specifically[75]. It is possible that the lysis process of this particular kit, or another step in the protocol, has off target effects in host depletion or biased effects on parts of the bacterial community. It is also possible, on the other hand, that use of this kit enables a more accurate view due to a more thorough lysis. Currently, the only comparison between this kit and another method are the two samples using the DNeasy PowerSoil Pro kit which – for unknown reasons – had a somewhat larger (albeit still small) amount of bacterial 16S reads. More comparisons are necessary for better understanding.

The overlap between bacterial ASVs found in animal samples and bacterial ASVs found in seawater samples is high. While horizontal[83] and vertical[84] bacterial transmission has been shown in marine invertebrates, microorganism of the animal microbiome can also originate from surrounding seawater[15]. At the same time, all seawater control samples also had detectable amounts of animal 18S, more than half of all detected reads in some cases, meaning that there was environmental DNA from animals present. Environmental DNA, other than living organisms, consists of expelled cells and particles, and extracellular DNA, which can be bound to substrates or free-floating[79]. Shed cells could include traces of intracellular symbionts of animals, and the release of symbiont DNA into the environment can also not be excluded. Due to the low overall numbers, it is therefore not possible

to determine which ASVs shared between animal and seawater samples originate from animals, which from seawater, and which are present in both.

As the water medium of the animals in captivity is filtered and exchanged twice a week, including one change of tank, the source of bacteria in seawater is also uncertain. While surprising differences in the seawater between symbiotic and aposymbiotic animals were detected, I sampled only small amounts of seawater intended as a control and do not have enough material or sample replication to test this. It could, however, be an interesting venture for future studies.

4.5. Isolated bacteria

Four bacteria were isolated from symbiotic C-line *Waminoa* worms, and one from the corresponding seawater control. Three of the isolated bacteria were identified as the same bacterial strain. Thus, despite the difficulties presented with it here, amplicon sequencing has routinely resulted in higher numbers of ASVs than culturing was able to. It was also possible to obtain more sequences from the seawater control, to determine with higher certainty which bacteria are found in the animals themselves, and which are from the environment. Amplicon sequencing has been shown to provide a better overview of the microbiome[85] compared to culturing, and only <20% of environmental bacteria are estimated to be able to grow in defined growth media[86]. Therefore, culturing is not sufficient to describe the bacterial community of the animal host, but serves as an interesting comparison with different techniques, and these cultured bacteria strains may still be useful for future experiments involving *Waminoa*, as despite its often noted limitations, the culturing of bacteria in corals is still actively researched and opens avenues not available to an exclusively metagenomics based approach[16].

5. Conclusion and Outlook

The prokaryotic microbiome of the *Waminoa* flatworm has proven particularly challenging to characterize with amplicon sequencing. In this thesis, I have tested multiple methods for DNA extraction, amplification, and sequencing, and compared the results.

Samples extracted with the DNeasy PowerSoil Pro kit and the DNeasy PowerBiofilm kit had similar results: save for one outlier, no 16S sequences were detected in any symbiotic C-line *Waminoa* animal samples, resulting in not enough data for appropriate in-depth analysis.

In contrast, samples extracted with the QIAamp DNA Microbiome kit performed highly successfully in terms of host depletion, but had notably different compositions from any other method. At the current time, it is impossible to determine which method is closer to the true community, and it is possible that the QIAamp DNA Microbiome kit has off-target effects resulting in biases towards some members of the microbial community. This discrepancy has been observed in other comparisons[75] as well. The price point is relatively high, but the application is simple. Concluding from the evidence in this thesis, I therefore recommend combining the QIAamp DNA Microbiome kit with at least one other method, to control for bias in DNA extraction.

Within the scope of this thesis, it was not possible to create a working LNA and corresponding protocol to effectively block amplification of *Waminoa* host 18S DNA. LNA design is multifactored, and even small changes in the PCR program can have an impact on its functionality. For this reason, LNAs require a large amount of testing and optimizing. They may still be viable options, however, if the necessary time is invested and long-term use is desired. Once an LNA and a corresponding protocol are successfully established for a particular use case, the LNA is, compared to other methods, low in cost to synthesize and reorder.

Photosymbiosis has long been seen as a two-player interaction, but in reality, it is a complex system of the host, the photosymbiont, and the multipartite symbiosis with all other microorganisms present in the system. The growing crisis of coral bleaching requires a greater understanding of this relationship, and at the same time, non-coral systems remain underrepresented in this field. Therefore, development of reliable, replicable, best-practice methods to characterize the prokaryotic microbiomes of such photosymbiotic systems is essential for creating studies that can be successfully carried out, compared, and can have their findings applied in practice.

6. References

1. Ogunlana, M. V. *et al.* Waminoa brickneri n. sp. (Acoela: Acoelomorpha) associated with corals in the Red Sea. *Zootaxa* **1008**, 1–11 (2005).
2. Biondi, P., Masucci, G. D., Kunihiro, S. & Reimer, J. D. The distribution of reef-dwelling Waminoa flatworms in bays and on capes of Okinawa Island. *Mar. Biodivers.* **49**, 405–413 (2019).
3. Barneah, O. *et al.* Three party symbiosis: acoelomorph worms, corals and unicellular algal symbionts in Eilat (Red Sea). *Mar. Biol.* **151**, 1215–1223 (2007).
4. LaJeunesse, T. C. *et al.* Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts. *Curr. Biol.* **28**, R873–R875 (2018).
5. Barneah, O., Brickner, I., Hooge, M., Weis, V. M. & Benayahu, Y. First evidence of maternal transmission of algal endosymbionts at an oocyte stage in a triploblastic host, with observations on reproduction in Waminoa brickneri (Acoelomorpha). *Invertebr. Biol.* **126**, 113–119 (2007).
6. de Bary, A. Ueber Symbiose. *Tageblatt Versamml. Dtsch. Naturforscher Aerzte* **51**, 121–126 (1878).
7. Martin, B. & Schwab, E. Current Usage of Symbiosis and Associated Terminology. *Int. J. Biol.* **5**, 32–45 (2012).
8. Yellowlees, D., Rees, T. A. V. & Leggat, W. Metabolic interactions between algal symbionts and invertebrate hosts. *Plant Cell Environ.* **31**, 679–694 (2008).
9. Jr, G. D. S. & Lipps, J. H. Photosymbiosis: The Driving Force for Reef Success and Failure. *Paleontol. Soc. Pap.* **17**, 33–59 (2011).
10. Boilard, A. *et al.* Defining Coral Bleaching as a Microbial Dysbiosis within the Coral Holobiont. *Microorganisms* **8**, 1682 (2020).
11. Wilkinson, C. Global and local threats to coral reef functioning and existence: Review and predictions. *Mar. Freshw. Res. - MAR Freshw. RES* **50**, 867–878 (1999).
12. Douglas, A. E. Coral bleaching—how and why? *Mar. Pollut. Bull.* **46**, 385–392 (2003).
13. Souter, D. *et al. Status of Coral Reefs of the World: 2020*. (2020).

14. Bien, T., Hambleton, E. A., Dreisewerd, K. & Soltwisch, J. Molecular insights into symbiosis-mapping sterols in a marine flatworm-algae-system using high spatial resolution MALDI-2-MS imaging with ion mobility separation. *Anal. Bioanal. Chem.* **413**, 2767–2777 (2021).
15. Apprill, A. Marine Animal Microbiomes: Toward Understanding Host–Microbiome Interactions in a Changing Ocean. *Front. Mar. Sci.* **4**, 222 (2017).
16. Sweet, M. *et al.* Insights into the Cultured Bacterial Fraction of Corals. *mSystems* **6**, (2021).
17. Rohwer, F., Seguritan, V., Azam, F. & Knowlton, N. Rohwer F, Seguritan V, Azam F, Knowlton N.. Diversity and distribution of coral-associated bacteria. Mar Ecol Prog Ser 243: 1-10. *Mar. Ecol.-Prog. Ser. - MAR ECOL-PROGR SER* **243**, 1–10 (2002).
18. Neave, M. J., Michell, C. T., Apprill, A. & Voolstra, C. R. Endozoicomonas genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Sci. Rep.* **7**, 40579 (2017).
19. Webster, N. S. & Reusch, T. B. H. Microbial contributions to the persistence of coral reefs. *ISME J.* **11**, 2167–2174 (2017).
20. Voolstra, C. R. & Ziegler, M. Adapting with Microbial Help: Microbiome Flexibility Facilitates Rapid Responses to Environmental Change. *BioEssays* **42**, 2000004 (2020).
21. Peixoto, R. S., Rosado, P. M., Leite, D. C. de A., Rosado, A. S. & Bourne, D. G. Beneficial Microorganisms for Corals (BMC): Proposed Mechanisms for Coral Health and Resilience. *Front. Microbiol.* **8**, (2017).
22. Bathia, J. *et al.* Symbiotic Algae of Hydra viridissima Play a Key Role in Maintaining Homeostatic Bacterial Colonization. *Front. Microbiol.* **13**, (2022).
23. De Wolfe, T. J. & Wright, E. S. Multi-factorial examination of amplicon sequencing workflows from sample preparation to bioinformatic analysis. *BMC Microbiol.* **23**, 107 (2023).
24. Lema, N. K., Gemeda, M. T. & Woldesemayat, A. A. Recent Advances in Metagenomic Approaches, Applications, and Challenges. *Curr. Microbiol.* **80**, 347 (2023).
25. López-Aladid, R. *et al.* Determining the most accurate 16S rRNA hypervariable region for taxonomic identification from respiratory samples. *Sci. Rep.* **13**, 3974 (2023).
26. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).

27. Coughlan, L. M., Cotter, P. D., Hill, C. & Alvarez-Ordóñez, A. Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Front. Microbiol.* **6**, (2015).
28. Ketchum, R. N. *et al.* DNA Extraction Method Plays a Significant Role When Defining Bacterial Community Composition in the Marine Invertebrate Echinometra mathaei. *Front. Mar. Sci.* **5**, (2018).
29. Clarke, L. J., Soubrier, J., Weyrich, L. S. & Cooper, A. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Mol. Ecol. Resour.* **14**, 1160–1170 (2014).
30. Kelly, R. P., Shelton, A. O. & Gallego, R. Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies. *Sci. Rep.* **9**, 12133 (2019).
31. Marotz, C. A. *et al.* Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* **6**, 42 (2018).
32. Heravi, F. S., Zakrzewski, M., Vickery, K. & Hu, H. Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *J. Microbiol. Methods* **170**, 105856 (2020).
33. Ganda, E. *et al.* DNA Extraction and Host Depletion Methods Significantly Impact and Potentially Bias Bacterial Detection in a Biological Fluid. *mSystems* **6**, e00619-21 (2021).
34. Kloß, S. *et al.* Destruction-free procedure for the isolation of bacteria from sputum samples for Raman spectroscopic analysis. *Anal. Bioanal. Chem.* **407**, 8333–8341 (2015).
35. Zhang, P. *et al.* Facile syringe filter-enabled bacteria separation, enrichment, and buffer exchange for clinical isolation-free digital detection and characterization of bacterial pathogens in urine. *Analyst* **146**, 2475–2483 (2021).
36. Sajib, M. S. I. *et al.* Advances in Host Depletion and Pathogen Enrichment Methods for Rapid Sequencing–Based Diagnosis of Bloodstream Infection. *J. Mol. Diagn.* **26**, 741–753 (2024).
37. Henderson, G. *et al.* Effect of DNA Extraction Methods and Sampling Techniques on the Apparent Structure of Cow and Sheep Rumen Microbial Communities. *PLOS ONE* **8**, e74787 (2013).
38. Frau, A. *et al.* DNA extraction and amplicon production strategies deeply inf luence the outcome of gut mycobiome studies. *Sci. Rep.* **9**, 9328 (2019).

39. Ahannach, S. *et al.* Microbial enrichment and storage for metagenomics of vaginal, skin, and saliva samples. *iScience* **24**, 103306 (2021).
40. Yap, M. *et al.* Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome. *Sci. Rep.* **10**, 21665 (2020).
41. Charalampous, T. *et al.* Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat. Biotechnol.* **37**, 783–792 (2019).
42. Prout, J., Tian, M., Palladino, A., Wright, J. & Thompson, J. F. LNA blockers for improved amplification selectivity. *Sci. Rep.* **13**, 4858 (2023).
43. Tan, S. & Liu, H. Unravel the hidden protistan diversity: application of blocking primers to suppress PCR amplification of metazoan DNA. *Appl. Microbiol. Biotechnol.* **102**, 389–401 (2018).
44. Ballantyne, K. N., van Oorschot, R. A. H. & Mitchell, R. J. Locked nucleic acids in PCR primers increase sensitivity and performance. *Genomics* **91**, 301–305 (2008).
45. Kurreck, J., Wyszko, E., Gillen, C. & Erdmann, V. A. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.* **30**, 1911–1918 (2002).
46. Yang, H. *et al.* A tailored LNA clamping design principle: Efficient, economized, specific and ultrasensitive for the detection of point mutations. *Biotechnol. J.* **16**, e2100233 (2021).
47. Fouz, M. F. & Appella, D. H. PNA Clamping in Nucleic Acid Amplification Protocols to Detect Single Nucleotide Mutations Related to Cancer. *Molecules* **25**, 786 (2020).
48. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
49. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
50. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).

51. Apprill, A., McNally, S., Parsons, R. & Weber, L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* **75**, 129–137 (2015).
52. Pjevac, P. *et al.* An Economical and Flexible Dual Barcoding, Two-Step PCR Approach for Highly Multiplexed Amplicon Sequencing. *Front. Microbiol.* **12**, (2021).
53. Team, R. R: A language and environment for statistical computing. *MSOR Connect.* (2014).
54. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
55. Bushnell, B. BBMap. http://sourceforge.net/projects/bbmap/
56. Laros, J. Demultiplex: FASTA/FASTQ demultiplexer. github.com/jfjlaros/demultiplex
57. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
58. Oksanen, J. *et al.* vegan: Community Ecology Package. (2024).
59. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
60. Coombes, K. R., Brock, G., Abrams, Z. B. & Abruzzo, L. V. Polychrome: Creating and Assessing Qualitative Palettes with Many Colors. *J. Stat. Softw.* **90**, 1–23 (2019).
61. Wu, Y.-H. *et al.* Brevirhabdus pacifica gen. nov., sp. nov., isolated from deep-sea sediment in a hydrothermal vent field. *Int. J. Syst. Evol. Microbiol.* **65**, 3645–3651 (2015).
62. Yakimov, M. M. *et al.* Alcanivorax borkumensis gen. nov., sp. nov., a new, hydrocarbon-degrading and surfactant-producing marine bacterium. *Int. J. Syst. Evol. Microbiol.* **48**, 339–348 (1998).
63. Sorokin, D. Y. Sulfitobacter pontiacus gen. nov., sp. nov. - A new heterotrophic bacterium from the Black Sea, specialized on sulfite oxidation. *Microbiology* **64**, 295–305 (1995).
64. Reimer, L. C. *et al.* BacDive in 2022: the knowledge base for standardized bacterial and archaeal data. *Nucleic Acids Res.* **50**, D741–D746 (2022).
65. Bookstaver, M., Godfrin, M. P., Bose, A. & Tripathi, A. An insight into the growth of *Alcanivorax borkumensis* under different inoculation conditions. *J. Pet. Sci. Eng.* **129**, 153–158 (2015).

66. Beiralas, R., Ozer, N. & Segev, E. Abundant Sulfitobacter marine bacteria protect Emiliania huxleyi algae from pathogenic bacteria. *ISME Commun.* **3**, 100 (2023).
67. Feng, T., Jeong, S. E., Kim, K. H., Park, H. Y. & Jeon, C. O. Aestuariicoccus marinus gen. nov., sp. nov., isolated from sea-tidal flat sediment. *Int. J. Syst. Evol. Microbiol.* **68**, 260–265 (2018).
68. Gavriilidou, A. *et al.* Comparative genomic analysis of Flavobacteriaceae: insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis. *BMC Genomics* **21**, 569 (2020).
69. Li, G. *et al.* Maricoccus atlantica gen. nov. sp. nov., isolated from deep sea sediment of the Atlantic Ocean. *Antonie Van Leeuwenhoek* **104**, 1073–1081 (2013).
70. WoRMS Editorial Board. World Register of Marine Species. https://www.marinespecies.org doi:10.14284/170.
71. López-Pérez, M. & Rodriguez-Valera, F. Pangenome Evolution in the Marine Bacterium Alteromonas. *Genome Biol. Evol.* **8**, 1556–1570 (2016).
72. Paul, S. I., Rahman, Md. M., Salam, M. A., Khan, Md. A. R. & Islam, Md. T. Identification of marine sponge-associated bacteria of the Saint Martin's island of the Bay of Bengal emphasizing on the prevention of motile *Aeromonas* septicemia in *Labeo rohita*. *Aquaculture* **545**, 737156 (2021).
73. Noguchi, T. *et al.* Vibrio alginolyticus, a tetrodotoxin-producing bacterium, in the intestines of the fish Fugu vermicularis vermicularis. *Mar. Biol.* **94**, 625–630 (1987).
74. Khan, A. A., Linkous, B. K. & Lanza, J. T. Vibrio alginolyticus: A Rare Cause of Otitis Externa off the Coast of Florida. *Cureus* **16**, e61524.
75. Eckensperger, S. Comparing host DNA depletion methods for seagrass leaf microbiome analysis. (Vienna, 2024).
76. Proctor, L. *et al.* A review of 10 years of human microbiome research activities at the US National Institutes of Health, Fiscal Years 2007-2016. *Microbiome* **7**, 31 (2019).
77. Boers, S. A., Jansen, R. & Hays, J. P. Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. *Eur. J. Clin. Microbiol. Infect. Dis.* **38**, 1059–1070 (2019).

78. Kalle, E., Kubista, M. & Rensing, C. Multi-template polymerase chain reaction. *Biomol. Detect. Quantif.* **2**, 11–29 (2014).
79. Scriver, M. *et al.* Drop it all: extraction-free detection of targeted marine species through optimized direct droplet digital PCR. *PeerJ* **12**, e16969 (2024).
80. Ikenaga, M. & Sakai, M. Application of Locked Nucleic Acid (LNA) Oligonucleotide–PCR Clamping Technique to Selectively PCR Amplify the SSU rRNA Genes of Bacteria in Investigating the Plant-Associated Community Structures. *Microbes Environ.* **29**, 286–295 (2014).
81. Bahl, M. I., Bergström, A. & Licht, T. R. Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis. *FEMS Microbiol. Lett.* **329**, 193–197 (2012).
82. Willis, A. D. Rarefaction, Alpha Diversity, and Statistics. *Front. Microbiol.* **10**, (2019).
83. Nussbaumer, A. D., Fisher, C. R. & Bright, M. Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* **441**, 345–348 (2006).
84. Sharp, K. H., Eam, B., Faulkner, D. J. & Haygood, M. G. Vertical Transmission of Diverse Microbes in the Tropical Sponge Corticium sp. *Appl. Environ. Microbiol.* **73**, 622–629 (2007).
85. Gupta, S. *et al.* Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Commun. Biol.* **2**, 1–7 (2019).
86. Hiergeist, A., Gläsner, J., Reischl, U. & Gessner, A. Analyses of Intestinal Microbiota: Culture versus Sequencing. *ILAR J.* **56**, 228–240 (2015).

7. Appendix

Table 1: List of all extractions with animal line, number of animals used, kit used, concentration, device used to measure concentration, and whether they have been sequenced. Sw. = Seawater, OOR = out of range

*Concentration not measured, but method discarded due to PCR results

Line	n	Kit	Conc. [ng/ul]	Device	Sequenced?
Sw. C		ZymoBIOMICS DNA Miniprep Kit	2.75	NanoDrop	No
C	5	ZymoBIOMICS DNA Miniprep Kit	37.73	NanoDrop	No
C	15	ZymoBIOMICS DNA Miniprep Kit	61.74	NanoDrop	No
C	5	ZymoBIOMICS DNA Miniprep Kit add.: homogenized, centrifuge (4200 x g) for 3 minutes	3.27	NanoDrop	No
C	15	ZymoBIOMICS DNA Miniprep Kit add.: homogenized, centrifuge (4200 x g) for 3 minutes	3.18	NanoDrop	No
Sw. C		DNeasy PowerBiofilm Kit	2.14	NanoDrop	No
C	5	DNeasy PowerBiofilm Kit	16.41	NanoDrop	No
C	15	DNeasy PowerBiofilm Kit	77.97	NanoDrop	No
C	5	DNeasy PowerBiofilm Kit add.: homogenized, centrifuge (24 x g) for 2 minutes	4.93	NanoDrop	No
C	5	DNeasy PowerBiofilm Kit add.: homogenized, centrifuge (94 x g) for 1 minutes	4.29	NanoDrop	No
C	5	DNeasy PowerBiofilm Kit add.: homogenized, centrifuge (94 x g) for 2 minutes	4.02	NanoDrop	No
C	5	DNeasy PowerBiofilm Kit	3.08	NanoDrop	No
C	15	DNeasy PowerBiofilm Kit	2.26	NanoDrop	No
C	5	DNeasy PowerBiofilm Kit	*		No
C	15	DNeasy PowerBiofilm Kit	*		No
Sw. C		DNeasy PowerBiofilm Kit	5.38	Qubit	Yes
C	10	DNeasy PowerBiofilm Kit	57.4	Qubit	Yes
C	10	DNeasy PowerBiofilm Kit	31.3	Qubit	Yes
C	10	DNeasy PowerBiofilm Kit	31.4	Qubit	Yes
Sw. Apo		DNeasy PowerBiofilm Kit	1.1	Qubit	Yes
Apo	10	DNeasy PowerBiofilm Kit	41.2	Qubit	Yes
Apo	10	DNeasy PowerBiofilm Kit	42.6	Qubit	Yes
Apo	10	DNeasy PowerBiofilm Kit	40.8	Qubit	Yes
Blank		DNeasy PowerBiofilm Kit	OOR low	Qubit	Yes
Apo	10	DNeasy PowerSoil Pro Kit	80.4	Qubit	Yes
Apo	10	DNeasy PowerSoil Pro Kit	70.2	Qubit	Yes
Apo	10	DNeasy PowerSoil Pro Kit	70.8	Qubit	Yes
Sw. Apo		DNeasy PowerSoil Pro Kit	0.2	Qubit	Yes
C	10	DNeasy PowerSoil Pro Kit	65.2	Qubit	Yes
C	10	DNeasy PowerSoil Pro Kit	92	Qubit	Yes
C	10	DNeasy PowerSoil Pro Kit	71.2	Qubit	Yes
Sw. C		DNeasy PowerSoil Pro Kit	1.5	Qubit	Yes

A	10	DNeasy PowerSoil Pro Kit	64.8	Qubit	Yes
A	10	DNeasy PowerSoil Pro Kit	50	Qubit	Yes
A	10	DNeasy PowerSoil Pro Kit	61	Qubit	Yes
Sw. A		DNeasy PowerSoil Pro Kit	0.27	Qubit	Yes
Blank		DNeasy PowerSoil Pro Kit	OOR low	Qubit	Yes
Blank		DNeasy PowerSoil Pro Kit	OOR low	Qubit	No
C	6	QIAamp DNA Microbiome Kit	0.249	Qubit	No
C	6	QIAamp DNA Microbiome Kit	0.315	Qubit	No
C	6	QIAamp DNA Microbiome Kit	0.316	Qubit	No
C	6	QIAamp DNA Microbiome Kit	0.258	Qubit	No
C	6	DNeasy Blood and Tissue Kit	31.1	Qubit	No
C	6	DNeasy Blood and Tissue Kit	34.8	Qubit	No
C	6	DNeasy Blood and Tissue Kit	28.5	Qubit	No
C	6	DNeasy Blood and Tissue Kit	25.8	Qubit	No
C	6	DNeasy PowerSoil Pro Kit	65.2	Qubit	No
C	6	DNeasy PowerSoil Pro Kit	67.4	Qubit	No
C	6	DNeasy PowerSoil Pro Kit	71.5	Qubit	No
C	6	DNeasy PowerSoil Pro Kit	55.3	Qubit	No
C	6	HostZERO Microbial DNA Kit	OOR low	Qubit	No
C	6	HostZERO Microbial DNA Kit	OOR low	Qubit	No
C	6	HostZERO Microbial DNA Kit	OOR low	Qubit	No
C	6	HostZERO Microbial DNA Kit	OOR low	Qubit	No
C	10	QIAamp DNA Microbiome Kit	0.165	Qubit	Yes
C	10	QIAamp DNA Microbiome Kit	0.225	Qubit	Yes
C	10	QIAamp DNA Microbiome Kit	0.291	Qubit	Yes
Sw. C		QIAamp DNA Microbiome Kit	OOR low	Qubit	Yes
Apo	10	QIAamp DNA Microbiome Kit	0.114	Qubit	Yes
Apo	10	QIAamp DNA Microbiome Kit	0.181	Qubit	Yes
Apo	10	QIAamp DNA Microbiome Kit	0.136	Qubit	Yes
Sw. Apo		QIAamp DNA Microbiome Kit	OOR low	Qubit	Yes
Blank		QIAamp DNA Microbiome Kit	OOR low	Qubit	Yes