National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

UNIVERSITY OF Southampton

UNIVERSITY OF SOUTHAMPTON

Faculty Of Natural And Environmental Sciences

School of Ocean and Earth Science

Microbiota of dominant Atlantic copepods: *Pleuromamma* sp. as a host to a betaproteobacterial symbiont

by

Sara Joan Javornik Cregeen

Thesis for the degree of Doctor of Philosophy

March 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Ocean and Earth Sciences

Thesis for the degree of Doctor of Philosophy

MICROBIOTA OF DOMINANT ATLANTIC COPEPODS: *PLEUROMAMMA* SP. AS A HOST TO A BETAPROTEOBACTERIAL SYMBIONT

by Sara Joan Javornik Cregeen

Copepods are the most abundant zooplankton group in the ocean and play a pivotal role as grazers of microorganisms and prey for larger animals. Furthermore, they are major contributors to the pool of dissolved organic material in the pelagic and therefore play an important role in the microbial loop. Although biology of copepods has been under investigation for more than a century, few studies have looked at the relationship between copepods and their associated bacteria. Could copepods be perceived as distinctive microbial hotspots in nutrient poor pelagic environment?

The microbiota of three *Pleuromamma* species, an abundant genus of copepods that migrate vertically from surface waters to several hundred meters water depth and back on a daily basis was investigated using various molecular and morphological techniques. The focus was on the differences in the bacterial community composition of these copepods along the Atlantic Meridional Transect, which traverses major oceanic biomes such as the subtropical gyres and the equatorial convergence region. Additionally, the community structure and stable isotope composition of the likely microplankton food source as well as the *Pleuromamma* copepods was assessed along the same cruise transect.

Sequencing of 16S rRNA tag libraries derived from individual *Pleruromamma* copepods showed a broad diversity of Bacteria associated with these copepods. These bacterial communities were uniform across the oceanic provinces and *Pleuromamma* species. These results were reflected in the uniformity of the prey community composition, however there was an indication that *Pleuromamma* copepods from different regions in the Atlantic rely on different food sources. Fluorescence *in situ* hybridisation and electron microscopy showed the presence of bacteria in the midgut region of the copepod guts and more specifically a high abundance of Betaproteobacteria.

The bacterial community of a dominant copepod species in the Atlantic has previously not been studied on such a large spatial scale and sample size. This study shows that the same bacterial taxa were associated with *Pleuromamma* copepods inhabiting distinct oceanic regions. Moreover, a betaproteobacterial genus not present in the water-column appears to be closely associated with *Pleuromamma.*

“I may not have gone where I intended to go, but I think I have ended up where I needed to be.” - Douglas Adams

Where Earth's tears flow,
Ubiquitous crustaceans
Paddle their oar-feet
- by Dr. Matthew Humphreys

Table of Contents

List of Tables

List of Figures

List of Appendices

DECLARATION OF AUTHORSHIP

I, **Sara Joan Javornik Cregeen**, declare that this thesis, and the work presented in it, are my own, and have been generated by me, as the result of my own original research.

Microbiota of dominant Atlantic copepods: *Pleuromamma* sp. as a host to a betaproteobacterial symbionts

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before initial submission.

Signed: 

Date: 10/11/2016

Acknowledgements

I have been extremely lucky to have had the opportunity to travel, meet and collaborate with many people over the past four years. This PhD has been a journey and would not have been possible without the contributions of all of them.

First of all I would like to thank my supervisor Mike Zubkov, who gave me the opportunity to pursue a project that allowed me to grow as a scientist and enter the wonderful world of symbiosis research. Thank you, Mike, for the continued support, constructive criticisms and encouragement when I needed it most. Thank you also to my supervisory panel, Tom Bibby, Duncan Purdie and Doug Connelly for helpful discussions and supervision of my progress. Another constant source of support during my PhD journey came from Manuela Hartmann, who was always there when I needed advice or just a chat. Thank you Manu!

I was lucky to have been able to join two AMT research cruises both of which were amazing experiences. Firstly, thank you to the crew and NMF technicians of the AMT22 (RRS James Cook) and AMT24 (RRS James Clark Ross) for all the technical support and high spirits on both cruises. Thank you also to the principle scientists: Glen Tarran and Tim Smyth, for making it such a smooth operation. Thanks also to Gay for helping me with handling the Micronet and keeping me company with some great chats! Finally, thank you to Pri, Sina, Liam, Greta, Monica, Ryan, Moritz, Cat, Nina and many others for making the long working hours enjoyable, being silly in our free time and creating and unforgettable cruise experience.

I would especially like to thank Erica Goetze for being a great mentor and thank you for letting me join your team, learn from all of you and collect my *Pleuromamma* samples on AMT24. I look forward to continuing our constructive collaboration in the future.

Much of this PhD wouldn't have been possible without Rudi Amann and Bernhard Fuchs who invited me to visit the MPI in Bremen. Initially to stay for a few months but it ended up being much longer. Thank you to all the Mollies and the Symbionts for instantly making me part of the group and encouraging me to take my project in directions I hadn't considered. All the tea kitchen discussions, office chats and group meetings made a huge difference. Greta, thanks for being part the 'tag story' with data, advice and chats. Clara and the Biogeos, thank you for running samples on the mass spec and contributing to my analysis. And Niko, I wouldn't have such beautiful EM pictures without your expertise. Thank you also to Nicole Dubilier for the scientific input and enabling me to take this project further in the next year.

Thank you to Manuela, Rhiannon and Niko for agreeing to read parts of my work and for all the useful insight and suggestions for improvement. A big thank you also to my Dad for proof reading my thesis.

I'm incredibly happy to have friends from all over the world and there are too many to thank on one page, but here goes. Carolyn, Leah and Tom – thanks for being the craziest, most supportive and wonderful housemates. Jen, Matthew, Rhi, James, Matt and Chris – my Southampton time would have been boring without you. Clara, Dima, Niko, Mario, Jime, David, Greta *et al.* – thanks for making me feel welcome and showing me the weird and wonderful world of Bremen. Katja, Grega, Eva, Dario, Barbara, Nika, Vesna, Tanja, Tea – thank you for always being available for a chat and for the unwavering friendship.

Adrien, thank you for being my rock, reminding me to have fun when all I could see was work and being a (science) geek with me. The past two years have been so much richer with you. Last but not least, thank you to my parents; both of you have encouraged me to be brave and leave my comfort zone, making my life so much more fun. Finally, thank you also to other friends and family for all the support along the years.

This project was funded by the NERC NOC core funding and the Max-Planck Institute for Marine Microbiology. Field collections on AMT24 were partially supported by ship-time funded under National Science Foundation (NSF) grant OCE-1338959 to E. Goetze (U Hawaii at Manoa).

Finally, thank you also to my two examiners, Phyllis Lam and David Pond, for a fruitful discussion during my viva and your suggestions for improving my thesis.

Abbreviations

13C	carbon-13
15N	nitrogen-15
AMT	Atlantic Meridional Transect
ANOVA	Analysis of Variance
Bet42	FISH probe targeting Betaproteobacteria
BLAST	Basic Local Alignment Search Tool
CFI	condition factor index
CHN-EA	Carbon-Hydrogen-Nitrogen Elemental Analyzer
COI	mitochondrial cytochrome oxidaes I gene
CTD	Conductivity-Temperature-Depth
DGGE	Denaturing Gradient Gel Electrophoresis
DOM	Dissolved Organic Material
EQ	Equatorial Region
EUB	general bacterial FISH probe
FISH	Fluorescence *in situ* hybridisation
Gam42	FISH probe targeting Gammaproteobacteria
MED	Minimum Entropy Decompostion
MPI-GC	Max Planck-Genome-centre (Cologne, Germany)
MPI-MM	Max Planck Institute for Marine Microbiology (Bremen, Germany)
NAG	North Atlantic Gyre
NanoSIMS	Nanoscale Secondary Ion Mass Spectrometry
NCBI	National Center for Biotechnology Information

NEC	North Equatorial Current
NMDS	Non-metric dimensional scaling
NT	Northern Temperate Region
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
PFA	paraformaldehyde
PL	prosome length
PLGP	*Pleuromamma gracilis* and *P. piseki*
PLBO	*Pleuromamma borealis*
PML	Plymouth Marine Laboratory
RDP	Ribosomal Database Project
RFLP	Restriction Fragment Length Polymorphism
SAG	South Atlantic Gyre
SDS	Sodium dodecyl sulfate
SEAc	Corrected Standard Ellipse Area
SEC	South Equatorial Current
SIA	Stable Isotope Analysis
SSU rRNA	small subunit ribosomal RNA
ST	Southern Temperate
TEF	Trophic enrichment factor
TEM	Transmission Electron Microscopy

Chapter 1

Introduction

The oceans cover approximately 71% of the Earth's surface and play an important role in determining the climate, for example, through heat storage and transfer, as well as being an environment that enabled life to evolve (Nybakken and Bertness, 2004). The marine pelagic environment of the ocean, defined as waters from the surface to the sea floor, is perhaps one of the largest biomes on Earth (Verity *et al*., 2002). It harbours a diverse group of organisms that are responsible for the photosynthetic fixation and conversion into biomass of approximately 50 Gt carbon per year (equivalent to half of the Earth's yearly carbon fixation) as well as the subsequent remineralisation and sequestration of this carbon to the sea floor (Angel, 1993; Field *et al*., 1998). The organisms and biological processes of the ocean have been termed the "biological pump", with the main driving force being the trophic relationships among phytoplankton, zooplankton and bacterioplankton (Longhurst and Harrison, 1989).

The Atlantic Ocean is characterised by distinct regions that are shaped by the physical and biological properties of the ocean such as ocean currents, temperature and salinity as well as chlorophyll a concentrations. The regions are most clearly differentiated by biomass abundance and nutrient availability (Maranon *et al*., 2000; Zubkov *et al*., 2000). Both oligotrophic gyres are characterised by low biomass and productivity, while temperate and equatorial (EQ) regions have higher abundances of primary producers and nutrient input (Moore *et al.,* 2013). Differences have been observed in the community composition and functioning of primary producers, as well as heterotrophic bacteria between the gyres and the EQ region, with organisms becoming very specialised to the specific limiting conditions of their environments. Single celled eukaryotes and cyanobacteria are the main primary producers in the open oceans (Jardillier *et al*., 2010). Abundances and the carbon biomass of the eukaryotic phytoplankton, *Synechococcus* cyanobacteria and heterotrophic bacteria are highest in regions with shallower picno- and nitra- clines, such as the EQ region (Tarran *et al*., 2006). Conversely, studies have found that *Prochlorococcus* cyanobacteria tend to dominate the biomass in the gyres (Zubkov *et al*., 1998) and have evolved distinct mechanisms that enable them to optimise uptake of essential organic nutrients, thus giving them a competitive advantage in a nutrient-depleted environment (Gomez-Pereira *et al.,* 2013; Zubkov *et al*., 2015). The gyres also

have lower abundances of heterotrophic bacteria than the EQ region (Zubkov *et al*. 1998, 2000; Heywood *et al.* 2006).

All regions, however, are capable of also supporting larger, metazoan plankton. Even in oligotrophic oceans, vertically migrating zooplankton such as copepods can be important consumers of the phytoplankton and phytodetritus (Lopez and Anadon, 2008) and their faecal pellets accelerate the vertical flux of carbon and nitrogen (Morales, 1999; Wilson and Steinberg, 2010). Dominant species of copepods form distinct populations within defined oceanic biomes (Goetze, 2011). This indicates that physical structuring of the ocean contributes to population differentiation of animals with supposedly wide dispersal ranges.

1.1 Copepods – pivotal players in the pelagic food web

The most abundant and widespread group of zooplankton in the ocean are copepods (Subphylum *Crustacea*; Class *Maxillopoda*; Subclass *Copepoda*). As a group they are extremely diverse in respect to external morphologies, habitat preferences and dietary modes. There are approximately 12,000 validated species living in marine and freshwater habitats, as well as hot springs, leaf-litter and even between sand grains. They are more abundant and widespread than any other metazoan group (Bron *et al*., 2011). The subclass consists of ten orders, with the order Calanoida being the most abundant of marine zooplankton (approximately 2000 species).

Although the order as a whole is taxonomically well defined, its taxonomy and phylogenetic relationships remain problematic. This is mainly due to the wide range of morphological characteristics, widespread and overlapping geographical ranges and large number of cryptic species complexes (Bucklin *et al.*, 2003; Blanco-Bercial *et al*., 2011). More effort has been made recently to gather large-scale molecular data using phylogenetic markers in order to create a database for the identification and phylogeny of this important group (Blanco-Bercial *et al*., 2014). The most common taxonomic molecular markers have been nuclear ribosomal genes (e.g. 18S ribosomal RNA; Bucklin *et al.*, 2003; Goetze, 2003), the internal transcribed spacer (ITS; Chu *et al*., 2001) and mitochondrial genes (e.g. encoding the cytochrome c oxidase subunits I and II; Folmer *et al*., 1994; Machida *et al*., 2004). These markers have been used in a variety of studies to elucidate cryptic species complexes (e.g. Halbert *et al*., 2013), population structure (Schizas *et al.*, 1999), molecular systematics within the Calanoida (e.g. Bucklin *et al*., 2003), species

diversity (e.g. Bucklin *et al*., 2010) and evolutionary relationships (e.g. Machida *et al*., 2005).

In the Atlantic Ocean, copepods can represent from 65%-99% of total metazooplankton abundance, depending on the size class. Copepod abundances vary significantly between oceanic regions (oligotrophic gyres vs. upwelling regions; Lopez and Anadon, 2008). Copepods are usually numerically dominated by small-sized species in tropical and subtropical oceanic regions and have been described as being well adapted to low food concentrations (Schnack-Schiel *et al.*, 2010). Smaller but extremely important and abundant species have traditionally been under sampled because the majority of studies using zooplankton nets have focused on species caught with larger mesh size nets (Gallienne and Robinson, 2001; Turner, 2004). Copepods can be found predominantly in the micro- (0.02-0.2 mm) and meso- (0.2-2.0 mm) zooplankton size classes but adult sizes vary from <0.1mm – 23cm (Bron *et al*., 2011).

The advance of molecular techniques has enabled a more detailed study of copepod feeding and also the study of *in situ* prey identification (Harper *et al.*, 2005; Njestgaard *et al.*, 2007; King *et al.*, 2008). Using Stable Isotope Analysis (SIA), it is possible to capture a range of complex interactions (e.g. flexible feeding modes), as well as track the energy and biomass flows in the ecosystem (Post, 2002; McCutchan *et al*., 2003). SIA is becoming a more and more common tool in tracing pathways of organic matter through food webs. Most copepods exhibit a vast range of dietary preferences (e.g. grazing, predation and parasitism), depending on the environment in which they live but are most accurately described as omnivores and opportunistic predators (Turner, 2004). They have been observed to degrade detrital material to various degrees through coprophagy (ingestion of pellets), coprorhexy (fragmentation of pellets) and coprochaly (loosening of pellets; Iversen and Poulsen, 2007). As sloppy feeders, copepods are continuously breaking up larger organic matter and making it more accessible to other organisms. Less is known about small species and larval (nauplii) feeding ecology (Lopez *et al*., 2007). Isotopic biomarkers, such as carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) stable isotopes, or lipid biomarkers, help to determine species-specific feeding or dietary preferences (e.g. Bell *et al.,* 2007; Pond and Ward, 2011; Teuber *et al*., 2014).

1.2 Pelagic trophic web

Zooplankton play a pivotal role in shaping pelagic ecosystems (Banse, 1995). Copepod's grazing and predation have an important effect on the control of phytoplankton and

microzooplankton populations. In turn copepods are predated upon by larger zooplankton and nekton affecting the size and recruitment success of commercially important fish stocks (Beaugrand *et al.,* 2003). This is a very simplified view of the pelagic ecosystem in the form of a chain, though (Sommer *et al.*, 2002).

The ecosystem is driven by much more complex direct as well as indirect interactions (Figure 1.1). Another level of the pelagic food web complexity was added by the discovery of the 'microbial loop' (Azam *et al.*, 1983). The microbial loop dominates oligotrophic waters in which smaller cells have a competitive advantage in nutrient acquisition (Fenchel, 2008). It is both a pathway for carbon fixation (by phototrophic prokaryotes) and organic matter degradation (heterotrophic bacteria). The microbial biomass is consumed by predatory nano- and micro- plankton that are predated upon by the next trophic level (e.g. copepods). Through the production of fecal pellets, copepods additionally contribute to the downward flux of organic material into the benthic environment (Verity and Smetacek, 1996; Frangoulis *et al.*, 2005) and, at the same time delay the sinking of fecal pellets and organic matter, via ingestion, thus giving them more time to be modified by microorganisms (Lampitt *et al.*, 1990; Møller and Nielsen, 2001).

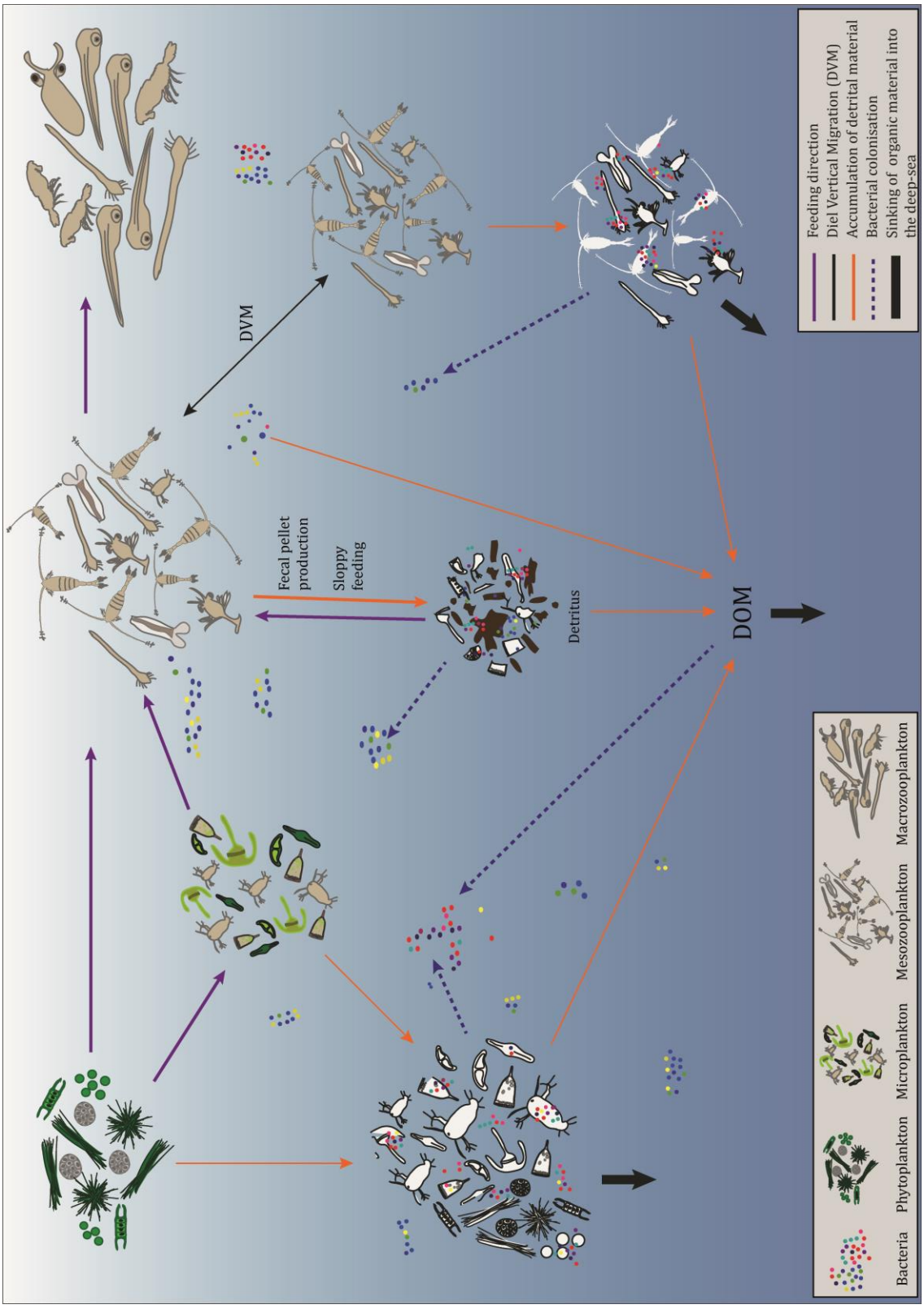


Figure 1.1 Summary of basic trophic interactions in the pelagic food web.

The main issue in assigning trophic levels to organisms and calculating the energy flow through the pelagic system is that most organisms have a nutritional flexibility, especially under low nutrient conditions (Sommer *et al.,* 2002). Many new trophic interactions, feeding mechanisms and functional groups are continuously being found. A novel discovery linked to the microbial loop is the mixotrophy of protists. Smallest plastidic protists or algae were found to be important bacterivores, as well as CO_2 fixers in oligotrophic oceanic gyres, indicating a need to change the way traditional food web functions are viewed (Zubkov and Tarran, 2008; Hartmann *et al*., 2012). Important but underestimated players in the microbial loop are viruses. Their ecological role is similar to grazers but often with higher host specificity. Viral infections can hence affect community structure as well as host population size (review by Fenchel, 2008 and references therein). Different zooplankton groups can either stimulate bacterioplankton growth by removing bacterivores or reduce bacterial abundance as their direct predators (Zubkov and Lopez-Urrutia, 2003). Zooplankton (e.g. copepods) can also stimulate bacterial growth by contributing to the pool of dissolved organic material (DOM) through a variety of feeding modes.

1.3 Animal-microbe associations

It is now widely accepted that most eukaryotes live in close association with prokaryotes. A long-term relationship between two or more organisms is defined as a symbiosis (from Greek *sym* 'with' and *biosis* 'living'). These relationships can range from a neutral association (commensalism) to a beneficial association for both partners (symbiosis, *s. str.*) or the benefit of one partner to the detriment of the other (parasitism). A hypothesis that species originated through close associations with microorganisms was first put forth by Mereschkowsky (*Mereschkowsky, 1910).* It gained cruicial support after molecular tools necessary for elucidating bacterial diversity and host-symbiont interactions were developed (Woese, 1987; Brucker and Bordenstein, 2012). It is a good example of how the fast development of sequencing and culture-independent molecular techniques has enabled ideas and theories proposed before their time to be proven and sometimes rediscovered. Decades later, Lynn Margulis' work confirmed that eukaryotic cells established a close obligate symbiosis early in evolutionary history with autotrophic and heterotrophic bacteria, which led to the development of organelles (plastids and mitochondria, respectively). These finidngs contributed to the symbiotic theory of evolution and laid the groundwork for research showing that microorganisms play an

integral role in the evolution and diversification of multicellular organisms (Sapp, 1994; Moya *et al*., 2008).

Furthermore, this has important implications for our understanding of the evolution of life, the on-going evolution of species, mechanisms of adaptation to various environmental pressures and interactions on the molecular level (Brucker and Bordenstein, 2012; McFall-Ngai *et al*., 2013; Kiers and West, 2015). It is becoming clear that organisms should not be viewed as individual units but rather as the sum of the host and all of their associated symbiotic microbiota, or even as the sum of all the genetic information of the host and the microbiota acting together as an individual. These ideas culminate in the hologenome theory of evolution, evidence for which is growing (Zilber-Rosenberg and Rosenberg, 2008; McFall-Ngai *et al*., 2013).

1.3.1 Types of symbioses

There are many different levels of host-microbe associations and the degree of dependency is related to the type of association (reviewed in Moya *et al*., 2008). There are two general types of symbiosis (Figure 1.2), based on the location of the symbionts. An ectosymbiosis is when symbionts are attached either to the body surface of the host, (epibionts) or internal surfaces of the host (e.g. digestive tract microbiota). Examples of an ectosymbiotic association are the filamentous Gamma- and Epsilon- proteobacteria living on specialised appendages of the *Rimicaris* hydrothermal vent shrimps (Petersen *et al.*, 2010). The stilbonematine nematods, for example, host *Cand.* Thiosymbion ectosymbionts on their body surface that are thought to provide a food source to the host (Ott *et al.*, 1991; Polz *et al.*, 1992). An endosymbiosis, on the other hand, is when the symbionts live within the host's cells. For example, hydrothermal mussels from the *Bathymodiolus* genus can host two different types of bacteria, a sulphur-oxidising chemoautotroph and a methane-oxidiser, within the cells of their gill tissue (Duperron *et al*., 2006). Another example of a deep-sea endosymbiosis are the vestimentiferan tubeworms hosting closely related gammaproteobacterial endosymbionts that have been hypothesised to allow these tubeworms to colonise a wide range of habitats (Zimmermann *et al.,* 2014). In some of the above examples, these symbioses enable both the animals and bacteria to survive in the chemosynthetic environment of hydrothermal vents and cold seeps.

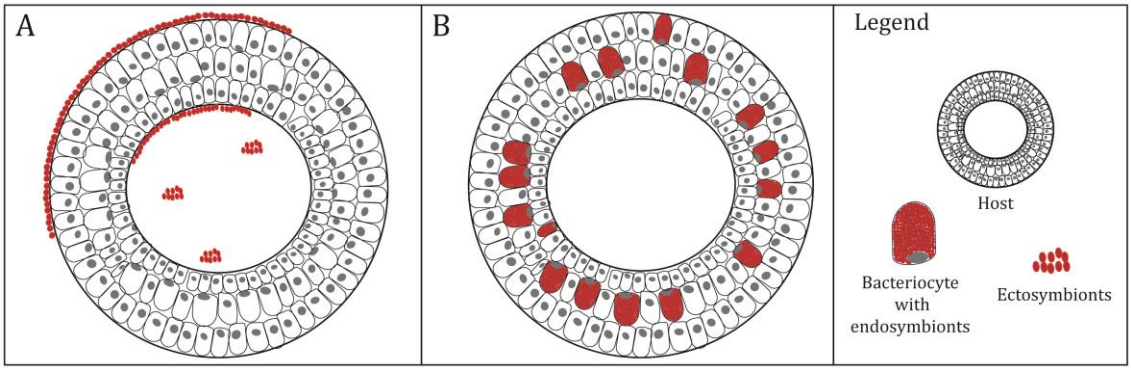


Figure 1.2 Basic types of symbiosis based on location of symbionts. **A**: Ectosymbiosis, where the symbionts are attached to the outside body surface or internal surfaces. **B**: Endosymbiosis, where the symbionts live inside the host's tissue.

Symbionts can further be either obligate (primary symbionts), often living in specialised host cells called bacteriocytes, or facultative (secondary symbionts). In some cases, the host can support both primary and secondary symbionts. One of the best-studied examples of such a symbiosis is in the pea-aphid host. They have an obligate primary endosymbiont *Buchnera*, which provides essential amino acids that the host cannot obtain through their diet of plant sap. Additionally, they have facultative secondary gammaproteobacterial symbionts, which provide a plant-specific fitness increase in the host's fecundity (Baumann *et al*., 1995; Tsuchida *et al*., 2004).

1.3.2 Symbiont role

The roles of symbionts in different systems depend on the service they provide for their hosts, for example defence against pathogens, nutrient acquisition, increased fecundity etc. The service the host provides its symbionts is usually a suitable habitat for colonisation, regular input of food or giving their symbionts access to both a source and sink of electrons for energy conversion (Chaston and Goodrich-Blair, 2010). Defensive symbioses are associations in which the symbiont protects its host from pathogenic infestations, for example the beewolf – *Streptomyces* symbiosis (Kaltenpoth *et al*., 2012). Symbionts can also positively affect the fecundity of their host, as is the case with the microbiota of the freshwater crustacean *Daphnia* sp. The majority of sequences obtained from fecund females belonged to *Limnohabitans* sp. (Betaproteobactera*)* and the authors successfully showed that the removal of the bacteria had a negative effect on the population size of the *Daphnia* (Peerakietkhajorn *et al*., 2015). In many symbiotic associations, the symbiont has a metabolic function and provides nutrition to its host. Chemoautotrophic symbioses are extremely widespread in the marine environment and

can be found in a wide range of habitats, such as deep-sea environments (e.g., hydrothermal vents, whale-falls and cold seeps) and shallow water sediments (Dubillier *et al.,* 2008). An alternative to a chemosynthetic symbiosis is a heterotrophic symbiosis, in which the symbiont does not transform chemical energy into organic carbon but rather helps its host with the digestion of nutritionally poor food that the host is unable to process (e.g. pea aphid-*Buchnera*, termite gut microbiota). In these associations, the symbionts usually colonise the host's digestive systems or specialised cells (Engel and Moran, 2013).

1.3.3 Symbiont transmission

There are two distinct ways by which a symbiont is transmitted from adult hosts to its progeny (Figure 1.3). They are important because the mode of transmission has implications on the diversity, phylogeny and evolution of the symbionts (Dubilier *et al*., 2008). Transmission can either be vertical – the symbionts are inherited directly from parent to offspring, usually through the maternal line, or horizontal – symbionts are acquired from the environment in early stages of host development, from a free-living symbiont source population (Bright and Bulgheresi, 2010).

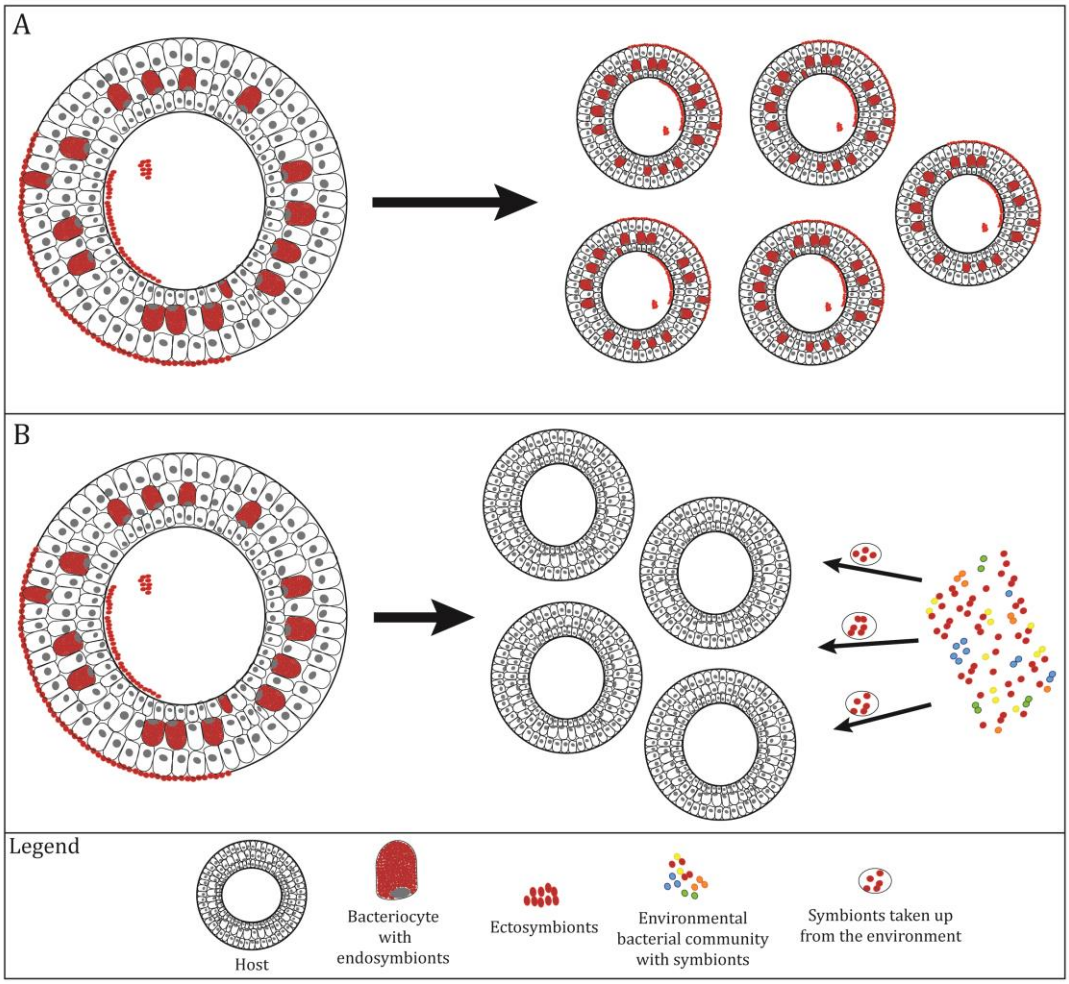


Figure 1.3 Two main modes of symbiont transmission. **A**: Vertical transmission. The symbionts are transmitted directly to the progeny from the parents. This can be the case for endo- and ecto-symbioses. **B**: Horizontal transmission. The offspring do not inherit the symbionts from their parents, but acquire them from the environment.

In cases of strict vertical transmission, symbiont populations experience a population bottleneck because only a small fraction of the bacterial partner is transferred to the next generation. Their population variability is thus very low, they have a reduced genome and, as a consequence exhibit the loss of a free-living lifestyle. However, the exact mode of transmission is still a mystery in many symbiosis systems and both modes are often present (Bright and Bulgheresi, 2010). An additional characteristic of strictly vertically transmitted symbioses is the parallel phylogenies and co-speciation of host and symbiont. This can be seen, for example, in the *Paracatenula* flatworms – thiotrophic Alphaproteobacteria symbiosis. By comparing host and symbiont phylogenies, Gruber-Vodicka *et al*. (2011) found that the symbiosis likely occurred in the very early evolution of the flatworms and has remained stable since. The host and symbiont coevolved, leading

to each host species having its own monoculture of the symbiont (Gruber-Vodicka *et al*., 2011).

Although vertical transmission is characteristic of well established, long lasting obligate symbioses, it is not necessarily the rule. In the squid light organ – *Vibrio fischeri* obligate symbiosis, the symbionts are transmitted horizontally between generations. Adult squid release their symbionts into the water column at dawn every day and their own population is then re-established from only a few individuals that are retained. The large numbers of *V. fischeri* bacteria now present in the water column are then available to hatchlings that are waiting to be colonised (Ruby *et al*., 1992; McFall-Ngai, 1999). Horizontal transmission is also characteristic of many gut bacteria associations, in which the acquisition of symbiotic bacteria is from the environment each generation. These associations can still be very specific and hosts have been shown to uptake the gut symbionts selectively (Engel and Moran, 2013).

1.4 Methodology used to study animal-bacteria interactions

Global prokaryotic diversity is extremely high and the true magnitude of it only started to be uncovered in the past 30 years. It is estimated that more than 90% of bacterial diversity is yet to be discovered. This includes diversity at the species level, as well as metabolic and functional diversity (Pedros-Alio, 2006). The concept of prokaryotic species has been reconsidered to accommodate the growing knowledge of diversity based on the sudden increase of sequencing. This was necessary because most new prokaryotes cannot be isolated in pure culture and described with the traditional morphological species concept (Rossello-Mora and Amann, 2001). Using of culture-dependant methods, a large part of the microbial diversity in both terrestrial and marine environments has been underestimated and perhaps a good indicator of this is that new organisms are still being discovered in systems that are considered low diversity, such as symbiotic associations (Dubilier, 2007). The field studying interactions between eukaryotes and prokaryotes has become increasingly important and widespread. Some of the main questions are the identity and abundance of the symbiont community, the specificity of the interaction, and the molecular mechanisms that enable the organisms to coexist.

The development of Sanger sequencing has allowed the characterisation of evolutionary relationships among microoragnisms and the small subunit ribosomal RNA (SSU rRNA) has become the most commonly used marker for microbial diversity (Woese, 1987; Pace, 1997). This has enabled a considerable diversity of microorganisms to be

described. The usefulness of rRNA gene sequences lies in the fact that all cells require ribosomes for translation of proteins needed for normal functioning; they are therefore present in high numbers and under low evolutionary pressure (Amann and Fuchs, 2008). More specifically, the 16S ribosomal (rRNA) gene has become the marker of choice used to assign phylogenetic affiliation to newly discovered prokaryotic organisms (Armougom and Raoult, 2009). The gene contains several hypervariable regions that allow differentiation of closely related organisms. Additionally, it has a low rate of horizontal gene transfer and recombination (Schloss *et al.*, 2011). Sanger sequencing has traditionally been used to characterise community diversity in combination with molecular fingerprinting techniques (e.g., restriction fragment length polymorphism – RFLP or denaturing gradient gel electrophoresis – DGGE) or clone libraries. The main drawback of traditional sequencing methods is that they do not allow high throughput analysis and the focus is on the most dominant taxa, so the rare taxa are often missed (van Elsas and Boersma, 2011).

1.4.1 Next generation sequencing technologies

Advances in high-throughput sequencing technologies have allowed a more in-depth and comprehensive analysis of the diversity of microorganisms in a variety of environments, including those associated with eukaryotes. The first next-generation sequencing platform to be developed was Roche 454 pyrosequencing (Margulies *et al.*, 2005), which revolutionised the field of studying microbial diversity. One of its first applications to marine microbial diversity used sequence tags from hypervariable regions to explore the phylogenetic diversity of deep water masses in the North Atlantic and they discovered a "rare biosphere" previously unknown (Sogin *et al.*, 2006). The technology has since been evolving at a very fast pace and many of the original approaches are becoming obsolete. Currently, the most used techniques are the Illumina sequencing platforms, which provide low-cost sequencing, increasing read lengths and paired-end reads (Schirmer *et al.*, 2016). However, an issue with these high-throughput sequencing is the high potential of error rates and this can lead to misinterpretation of the results (Schloss *et al.*, 2011; Loman *et al.,* 2012). All of the above sequencing methods have the issue of PCR introduced errors and bias. Other common errors can be chimera formation and errors introduced during sequencing, such as homopolymers and substitution type miscalls (Metzker, 2010; Schirmer *et al.,* 2015). A comparison of these sequencing methods is given in Table 1.1. Other sequencing platforms are also being developed that will produce longer read lengths, short sequencing time and low cost. These include for example the

nanopore-based sequencing devices (Branton *et al.*, 2008) and the Single-molecule real-time (SMRT) sequencing developed by Pacific BioScineces (PacBio). Although these 'third generetaion' sequencing technologies can provide longer reads and faster sequencing times, they have a lower throughput and are currently hindered by high error rates (Rhoads and Au, 2015).

Table 1.1 Comparison of leading next-generation sequencing technologies. (Shendure and Ji, 2008; Metzker, 2010; Loman *et al*., 2012).

Sequencing platform	Sequencing principle (Sequencing by synthesis)	Read length	Error type	Aditional comments
Illumina MiSeq HiSeq (2500)	Detect fluorescently labelled bases as they are incorporated	Paired-end reads 2 × 300 bp 2 × 150 bp	Substitution type miscalls	Currently the most used sequencing system on the market. e.g. Amplicon sequencing e.g. Genome and transcriptome sequencing
Roche 454 technology	Detection of pyrophosphate release	up to 600 bp	Homopolymers	Discontinued
Ion Torrent	Detection of hydrogen ions	Approx. 100 bp	Homopolymers	Fastest throughput and shortest run time

One of the most critical steps for accurate rRNA amplicon analysis is choosing appropriate PCR primer pairs. This can reduce PCR bias, such as preferential amplification of certain taxa leading to under-representation or even selection against individual groups (Klindworth *et al*., 2012). Other ways of reducing bias is using double indexing of primer sequences, which improves the accuracy of barcode splitting in the post-sequencing processing, while adding heterogeneity spacers improves sequencing quality (Fadrosh *et al.*, 2014). Databases containing 16S rRNA sequence information are growing rapidly and enabling the development of online services providing tools for testing primer pair specificity and overall coverage for target groups (e.g. ProbeCheck or TestPrime; Loy *et al.*, 2008; Klindworth *et al.*, 2012). By using appropriate primer pairs and appropriate strategies of downstream analysis of amplicons depending on the sequencing error type, misinterpretation of results can be greatly reduced. However, an extra caveat should be considered when using primers containing extra barcode sequences. It has been shown

that these barcoded primers can introduce a bias to the PCR and influence the reproducibility of the high-throughput sequencing datasets (Zhou *et al.*, 2011). This could be avoided by using techniques similar to 'reconditioning PCR', where the template is amplified without the barcoded primers in a first step PCR and the barcodes are added in a second step with lower cycle numbers (Berry *et al.*, 2011).

1.4.2 Partitioning of sequences into meaningful units

The traditional way of partitioning a large number of sequences into meaningful operational taxonomic units (OTUs), prior to taxonomic path assignment, has been the clustering approach. OTU clustering (Figure 1.4B) primarily relies on computing distance matrices from pairwise alignments to group sequences at a user specified threshold (typically >= 97%) and outputting a representative sequence that can be fed into taxonomic identification pipelines (Seguratin and Rohwer, 2001). Several new approaches have been developed since, including Oligotyping and Minimum Entropy Decomposition (MED) approaches (Eren *et al.*, 2014 a,b). Both approaches use the differences in Shannon entropy among sequence reads iteratively to decompose a dataset until the maximum entropy is reached (i.e. sequences are partitioned into groups that have no more entropy to explain; Figure 1.4A). This allows for a higher resolution and the partitioning of sequences into phylogenetically homogenous units ("MED nodes") that can show hidden diversity in closely related organisms to the nucleotide level (Eren *et al.*, 2014b). Characterisation of a bacterial community in a sequenced sample is only possible when taxonomic information can be assigned to sequences. There is thus a need for reliable, curated databases, which are regularly updated and improved. Several such online tools are available, for example, the Basic local alignment search tool (BLAST; Altschul *et al.,* 1990), SILVA (Quast *et al.*, 2013) or Ribosomal Database Project (RDP; Cole *et al.,* 2014).

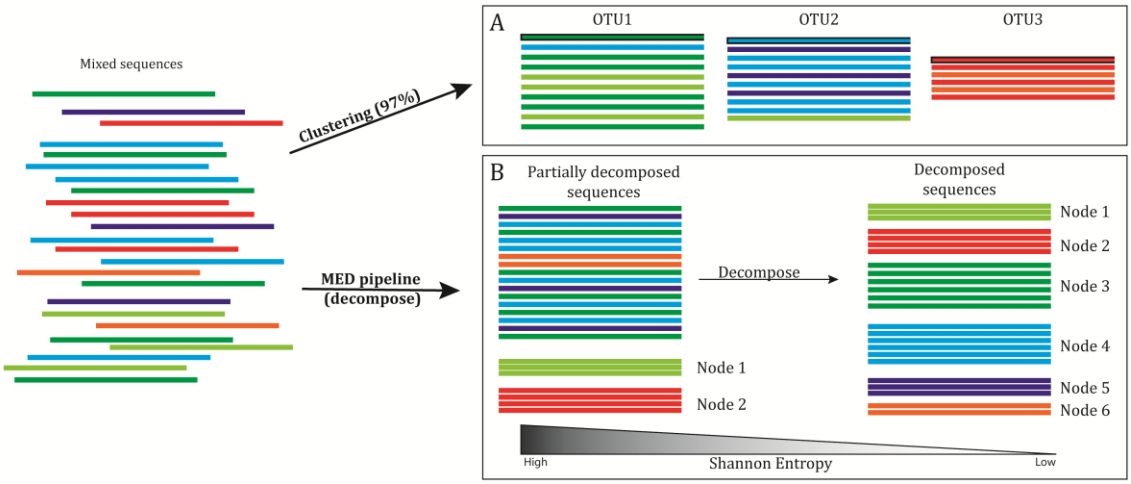


Figure 1.4 Two methods for partitioning a mix of sequences into meaningful taxonomic units. **A:** Traditional clustering approach. Sequences are aligned to reference sequences (black border) at a user defined threshold (usually 97%). **B:** Minimum Entropy Decomposition approach (MED) uses the differences Shannon Entropy to decompose a group of mixed sequences by iteratively grouping identical sequences into MED Nodes. The process is complete when the lowest Entropy is reached.

1.4.3 Quantitative and visualisation methods

Although high-throughput sequencing has provided new insight into the distribution and diversity of prokaryotic organisms worldwide, it lacks the reliable quantitative information that can be obtained using other methods. Quantification can be done with a variety of methods, such as for example DAPI staining or quantitative PCR. A method that allows quantification, phylogenetic identification and visualisation of microbial communities is *in situ* is Fluorescence *In Situ* Hybridisation (FISH; Amann *et al.*, 1990). Since its development, the method has become widely used and has been combined with other techniques, such as flow cytometry, NanoSIMS, microautoradiography, etc., to answer a range of ecological questions (Amann and Fuchs, 2008). The power of the method is in the use of oligonucleotide probes targeting specific regions of the rRNA, in combination with a variety of fluorescent labels that allow both phylogenetic identification of an organism and visualisation of the community *in situ* (Amann *et al.,* 1995). Because of the evolutionary stable nature of the rRNA molecule and growing quality of RNA databases, oligonucleotide probes can be designed to target a wide range of taxonomic levels, from domains to orders (Amann and Fuchs, 2008).

Visualisation of bacteria is especially useful in symbiotic systems in which the spatial distribution as well as identification of a symbiont community can be achieved using FISH in combination with histological embedding in a variety of media (e.g. paraffin wax or

different resins). When using a comparative 16S rRNA and FISH approach, it is possible both to identify and locate even low abundance symbionts in various animal-bacteria associations, such as the deep-sea *Rimicaris* shrimp and *Bathymodiolus* mussels (Petersen *et al.,* 2010; Assie *et al.*, 2016). When epifluorescence microscopy is combined with other methods, such as high-resolution electron microscopy, labelled substrate and stable isotope uptake and mass spectrometry, inferences can be made about the potential role of a symbiotic association.

The fast development of sequencing and other molecular techniques targeting the SSU rRNA molecule have enabled a better understanding of the prokaryotic community in marine systems and especially the diversity of symbiotic associations. Sequencing technologies have also enabled ecological studies to go beyond rRNA sequencing and investigate the genomic and metabolic capabilities of microorganisms. Although a large database of microbial phylogenetic diversity based on ribosomal RNA (rRNA) sequences now exists, there is still a huge diversity of metabolic pathways and genomic information that remains unknown.

1.5 Thesis rationale

Copepods have an extremely important role in the pelagic ecosystem but it is still unclear whether they have a distinct and permanent association with potentially beneficial bacteria. Many animals require symbiotic bacteria, either for energy or for essential nutrient supply, so it seems likely that copepods, living in oligotrophic pelagic environments, may have developed similar associations. The genus *Pleuromamma* was selected because it is an important component of the pelagic zooplankton community in the Atlantic Ocean and samples were collected on two Atlantic Meridional Transect cruises, in 2012 and 2014 (AMT22 and AMT24, respectively). This transect is especially interesting because it traversed major oceanic provinces in the Atlantic Ocean, such as the subtropical gyres (NAG and SAG) and the Equatorial convergence (EQ) region. This provided a unique platform to study and compare *Pleruomamma*-associated microbota across a large spatial scale.

1.5.1 Aims and objectives

The overarching aim of this study was to identify and characterise the gut microbiome of dominant Atlantic open-ocean copepods from the genus *Pleuromamma*. In order to do so comprehensively, the following objectives were set:

1. **To determine the phylogenetic relationship and describe morphological characteristics of *Pleuromamma* copepods (Chapter 3):** The molecular characterisation and phylogenetic relationship of three species of *Pleuromamma* copepods was investigated using a mitochondrial gene marker (COI). Body sizes were measured and condition indices calculated for a suite of samples along the transect to examine the fitness of animals inhabiting different regions of the Atlantic Ocean. Lastly, the external and internal morphology of *P. gracilis* was studied. An emphasis was given to the gut morphology in order to determine where the gut microbiota might be contained.

2. **To determine the role of microplankton as a food source for *Pleuromamma* copepods (Chapter 4):** The stable carbon and nitrogen composition, as well as elemental content was measured for microplankton in the 20-100 µm and 100-180 µm size fractions, which represent are a potential food source of the *Pleuromamma* copepods. Samples were also analysed using FlowCam in order to determine the community composition of the microplankton size fractions along the transect. Additionally, the stable carbon and nitrogen composition of a small set of *Pleuromamma* copepods was measured to determine their trophic position. The hypothesis behind this objective was that the diet of the *Pleuromamma* copepods could be inferred based on the isotopic composition of the potential microplankton prey. Potential differences in the host's trophic ecology could then help explain observations made about the gut microbiome.

3. **To describe the diversity of bacteria associated with *P. gracilis/piseki* and *P. borealis* copepods (Chapter 5):** The community composition of *Pleuromamma*-associated bacteria was investigated at the ocean basin scale. Owing to the small size of the host species (approximately 1.5 mm in length), gut excision was impractical given time constraints. An Illumina sequencing approach was used to determine the community composition across a range of samples. The diversity among different biomes and between the host species was explored. The main hypothesis driving this objective was that differences among regions, in host species, and, consequently, diet would affect the community composition of the *Pleuromamma*-associated bacteria. In conjunction, a hypothesis that specific copepod-associated bacteria would differ from the bacteria inhabiting the water-column was tested.

4. **To characterize specific Betaproteobacteria associated with the gut of *Pleuromamma* copepods (Chapter 6):** The most interesting finding of Chapter 5 was the presence of a *Pleuromamma*-associated betaproteobacterial sequence that was identical along the whole transect. It was present in a majority of the sequenced copepod samples, but absent from the water-column. Therefore, the hypothesis that this bacterial taxon, related to the genus *Limnobacter,* would inhabit the copepod's gut was tested. The visualisation and spatial distribution of the *Pleuromamma* gut microbiome is described using different microscopy approaches. Additionally, the phylogenetic relationship of sequences related to the betaproteobacterial genus *Limnobacter* was assessed and a specific oligonucleotide probe targeting the genus was developed.

Chapter 2

Atlantic Meridional Transect – description of the study site and sampling procedures

2.1 Atlantic Meridional Transect Programme

The Atlantic Meridional Transect Programme (AMT) is an international, multidisciplinary research effort that has been undertaking a series of physical, chemical and biological measurements on biannual or annual research cruises for the past 20 years (Rees *et al.,* 2015). The research cruises were originally ones of opportunity taking place during the passage of the RRS James Clark Ross on its annual voyage from the UK to ports in South America, South Africa or the Falkland Islands in the boreal autumn and back to the UK in the boreal spring (Aiken *et al.*, 2000). Since 2008 the research cruises have only taken place once per year, between September and November from the UK to South America or the Falkland Islands. Since the beginning in 1995, there have been 25 AMT cruises. The main subjects and goals of the AMT Programme are (Rees *et al.,* 2015):

(1) Investigating ecological and biogeochemical variability of planktonic systems – ranging from bacterio- and picoplankton to phyto- and zooplankton.
(2) Adding to research on nutrient cycling, as well as the export and air-sea exchange of climate active gases on a large spatial scale and between ocean basins.
(3) Contributing to wider multidisciplinary ocean time-series, which provides a more detailed picture of oceanic processes.
(4) Providing data that help validate both current and next-generation satellite algorithms.
(5) Taking measurements that help develop and validate global ecosystem models.

The transect covers over 10,000 km and crosses many distinct ecosystems in the Atlantic Ocean, such as oligotrophic oceanic gyres, and tropical and upwelling systems. It is an ideal platform for elucidating the diversity and potential differences in associations among important members of the mesozooplankton and bacterioplankton community on a large spatial scale.

The AMT22 research cruise took place on the RRS James Cook between Southampton (UK) and Punta Arenas (Chile) from 10th October to 24th November 2012. The AMT24 research cruise took place on the RRS James Clark Ross between Immingham (UK) and Stanley (Falkland Islands, UK) from 22nd September to 3rd November 2014.

2.2 Physical and biological structuring of the Atlantic Ocean

Distinct regions can be observed in the Atlantic Ocean, which are defined by nutrient availability, physical structuring and ocean currents (Maranon et al., 2000). The largest two regions are the Northern (NAG) and Southern (SAG) oligotrophic gyres on either side of the tropical equatorial (EQ) region and flanked by the Northern Temperate (NT) and Southern Temperate (ST) regions, respectively. The gyres together cover approximately one third of the area of the Atlantic Ocean and are characterised by the limited availability of inorganic nutrients such as nitrogen and phosphorus as well as dissolved organic matter (Moore *et al.*, 2013). They also have a distinctly deep pycnocline and nitracline, and chlorophyll a concentrations below 0.2 mg m^{-3} (Maranon *et al.*, 2001). Recent time-series studies have also shown that the oligotrophic regions around the world are in fact becoming larger, increasing annually (Polovina *et* al., 2008; Irwin and Oliver, 2009). The EQ on the other hand has both a shallower pycnocline and a deep chlorophyll maximum (Maranon *et al.*, 2001).

These biological and chemical properties of the Atlantic Ocean (e.g. nutrient availability in the surface waters) are a consequence of ocean circulation. The North Equatorial Current (NEC) creates clockwise circulation of the NAG while the South Equatorial Current (SEC) creates anti-clockwise circulation in the SAG. These act to deepen the pycnocline and prevent nutrient mixing from deep waters into the euphotic zone (Niiler, 2001). In the EQ, the currents discussed above, create an oceanographic front that separates the North Atlantic Central Water from the South Atlantic Central Water. Additionally other smaller sub-surface currents, upwelling and coastal processes act to shape this ocean province by input of nutrient rich deep waters into the surface layer (Tomczak and Godfrey, 2003). Another important source of nutrients (such as iron and phosphorus) are dust depositions in the upper ocean, which can drive specific bloom events in otherwise nutrient poor waters (Calil *et al.*, 2011).

2.2.1 Defining Atlantic Ocean regions along AMT22 and AMT24

The oceanic region boundaries along the AMT22 and AMT24 cruise tracks were determined according to patterns observed by Zubkov *et al*. (1998), Zubkov *et al.* (2000), Maranon *et al.* (2001), Heywood *et al.* (2006) and Tarran *et al.* (2006). The regions were used in further analysis to separate samples along the transect into groups for analysis and comparison.

Synechococcus cell abundances were used to determine region boundaries (Zubkov *et al.,* 2000) based on low abundance in the gyres and higher abundance in equatorial and temperate waters (Figures 2.1A and 2.2A). Samples were collected and analysed onboard by Glen Tarran (PML). Briefly, surface water samples were collected from a Seabird CTD system deployed at the pre-dawn and mid-day stations. The phytoplankton community structure was characterized and enumerated using a Becton Dickinson FACSort flow cytometer (for more detail see AMT24 Cruise Report, p. 57).

Additionally, a proxy for chl a concentrations (fluorescence) and the physical structuring of the water-column were used to determine the region boundaries. Calibrated data files for 70 CTD casts obtained at the pre-dawn and mid-day stations were used to plot profiles for fluorescence , salinity and temperature (Figures 2.1B-D and 2.2B-D). The profiles show a 300 m depth profile along a 50° N to 50° S transect. Plots were produced using Ocean Data View software (version 4.7.0, R. Schlitzer).

Based on the above methods, the following region boundaries were identified along the AMT 22 and AMT24 cruise tracks and used to group stations for statistical analysis:

		AMT22	AMT24
(i)	North Atlantic Gyre (NAG)	17° – 31 ° North	17° – 33 ° North
(ii)	Equatorial Region (EQ)	7° South – 16 ° North	9° South – 16 ° North
(iii)	Southern Atlantic Gyre (SAG)	8° – 28° South	10° – 31° South

This study did not focus on samples from the Northern or Southern Temperate waters.

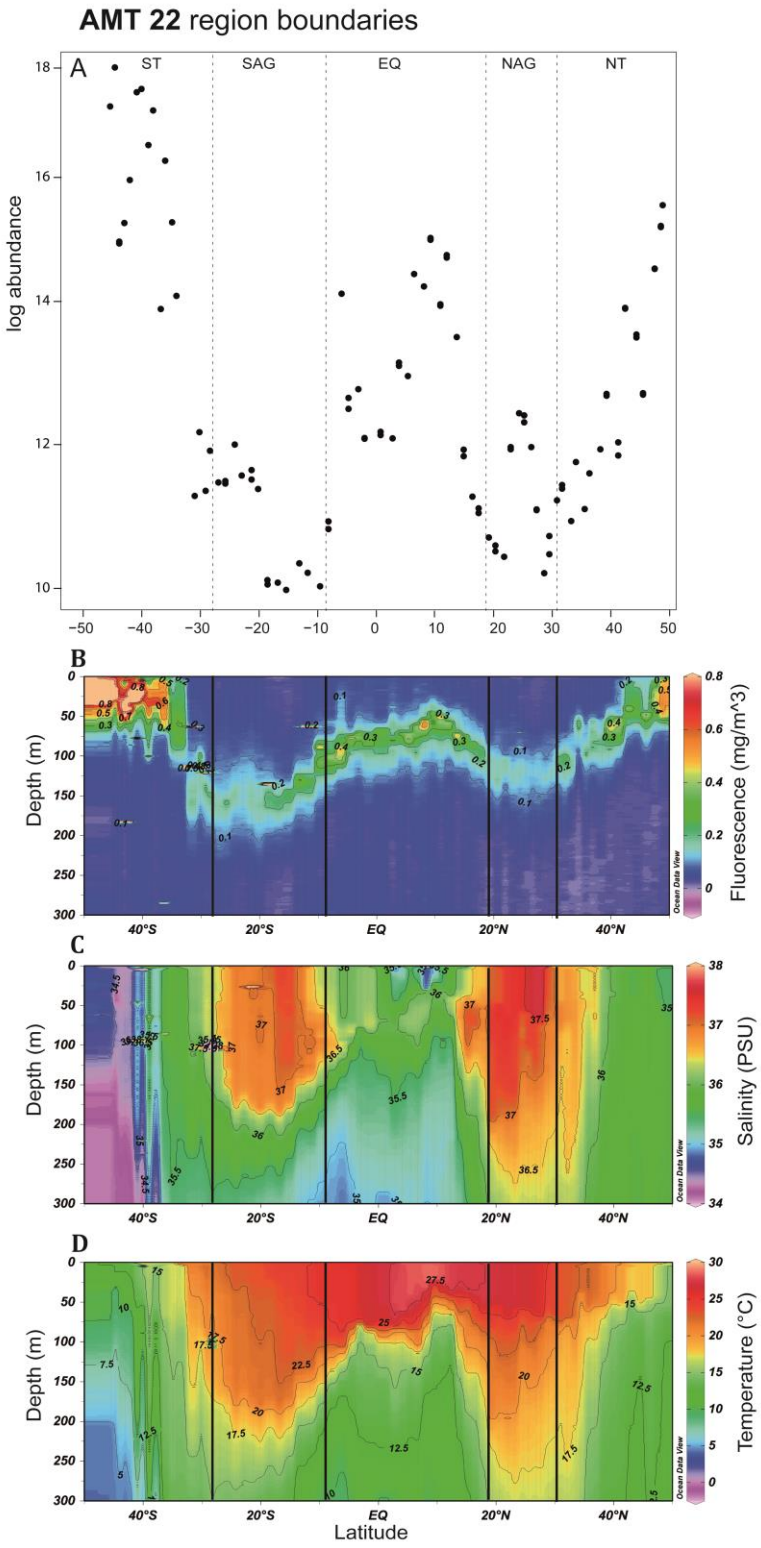


Figure 2.1 Region boundaries along the AMT22 cruise track. **A**: Surface *Synechococcus* cell abundances (log scale). (Data collected by G. Tarran, PML). **B-D**: CTD 300 m depth profiles showing Fluorescence (B), Salinity (C) and Temperature (D). Lines show region boundaries. From south to north: ST – Southern Temperate, SAG – South Atlantic Gyre, EQ – Equatorial Region, NAG – North Atlantic Gyre and NT – Northern Temperate Region.

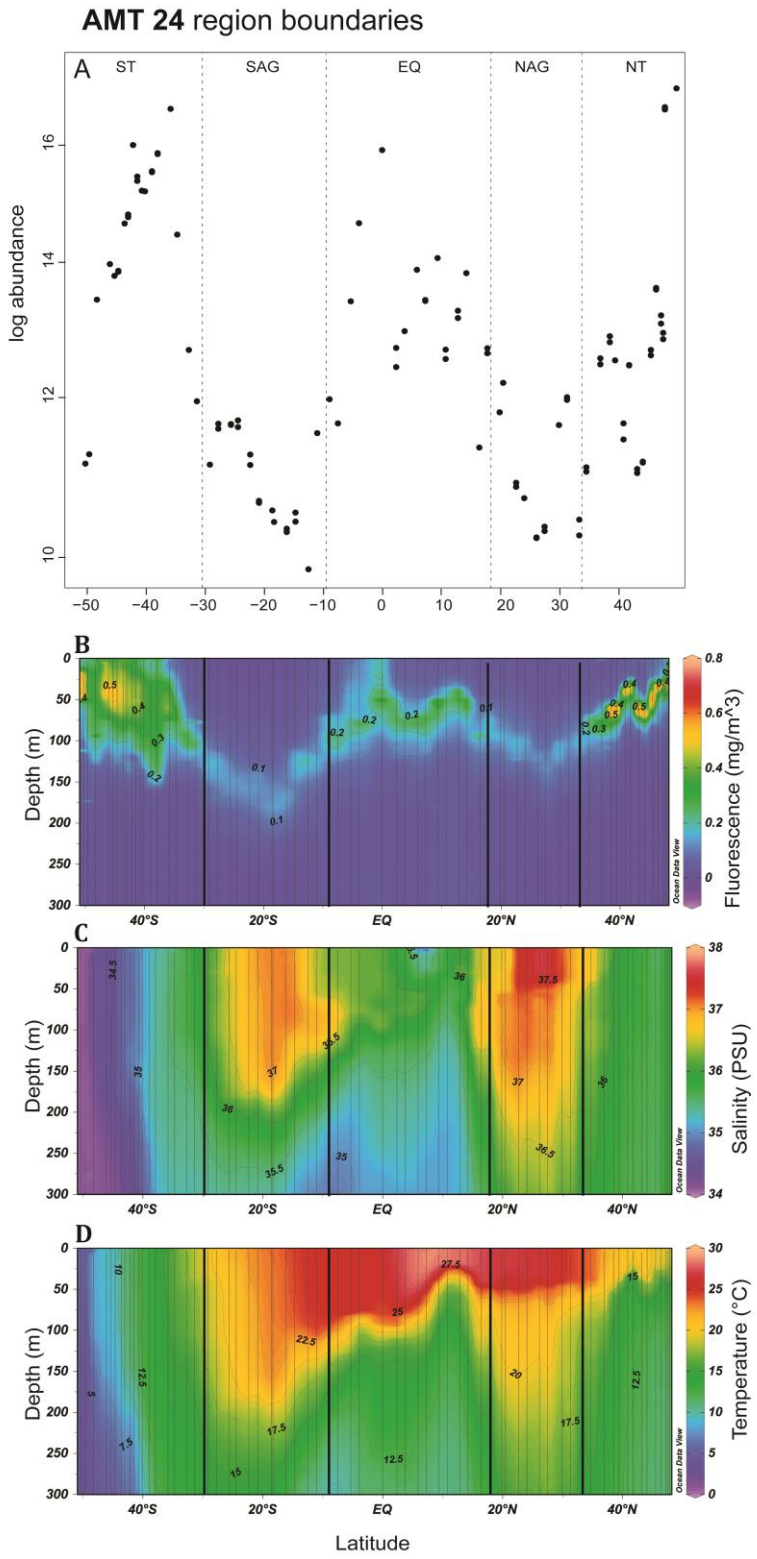


Figure 2.2: Region boundaries along the AMT24 cruise track. **A**: Surface *Synechococcus* cell abundances (log scale). (Data collected by G. Tarran, PML). **B-D**: CTD 300 m depth profiles showing Fluorescence (B), Salinity (C) and Temperature (D). Lines show region boundaries. From south to north: ST – Southern Temperate, SAG – South Atlantic Gyre, EQ – Equatorial Region, NAG – North Atlantic Gyre and NT – Northern Temperate Region.

2.3 AMT – sample collection

AMT research cruises follow a uniform sampling schedule, with two main sampling stations per day. The pre-dawn stations enabled sampling of communities and biological processes that are affected by daylight (e.g. migratory zooplankton) and mid-day stations allow measurements and sampling of contrasting conditions as well as optical measurements. Table 2.1 highlights the main sampling activities that contributed to this study. Figure 2.2 shows examples of the plankton sampling nets. The following sections describe the sampling on AMT22 and AMT24 in more detail.

Table 2.1 Sampling activity on AMT22 and AMT24 that contributed to this study

Sampling station	Time of day (ship time)	Cruise	Station code	Activity
Horizontal net tow	Pre-dawn station (approx. 23:00-23:30, selected days; and approx. 3:00-4:00 every day)	AMT24	ZPT	Collection of zooplankton material for live sorting
Vertical net tow	Pre-dawn station (4:00-5:00 every day)	AMT22	ZPV	Bulk zooplankton collection
CTD cast	Pre-dawn station (approx. 4:00-5:00); Mid-day station (approx. 13:00-14:00)	AMT22 and AMT24	CTD	Water-column sampling and measurements
Microplankton sampling	Mid-day station (approx. 14:00-15:00)	AMT22 and AMT24	MPV	Collection of microplankton material for stable isotope analysis and FlowCam

Figure 2.3 Nets used for pre-dawn zooplankton (**A**) and mid-day microplankton (**B**) sampling on AMT24 cruise.

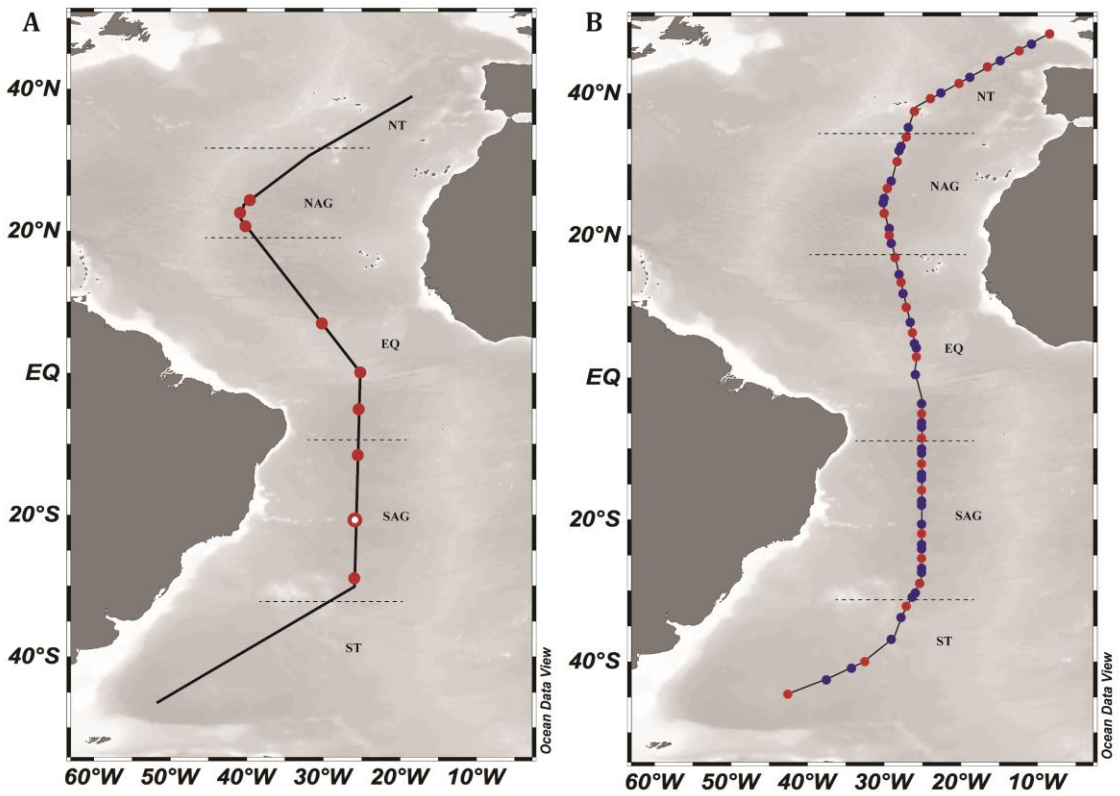


Figure 2.4 Map showing AMT22 (A) and AMT24 (B) cruise tracks with locations of stations used to collect samples used in this study. **A**: Stations at which individuals for Illumina sequencing (red) and the station at which the individual used for clone library construction was collected (white) on AMT22. **B**: AMT24 collection stations. Blue: pre-dawn stations (ZPT and CTD) Red: Mid-day stations (MPV and CTD). Dashed lines show region boundaries (NT – Northern Temperate, NAG – North Atlantic Gyre, EQ – Equatorial Region, SAG – South Atlantic Gyre and ST – Southern Temperate Region).

2.3.1 AMT22

Samples were collected at the pre-dawn stations of the 2012 Atlantic Meridional Transect Cruise (AMT22). Bulk zooplankton samples were collected at 29 pre-dawn stations from 200 m vertical net hauls using a double (bongo) net frame with 0.57 m diameter openings and WP2 nets with 200 µm nylon mesh. These were collected as part of a separate study of zooplankton community size structure by C. Gallienne from Plymouth Marine Laboratory (see AMT22 Cruise Report). The samples not needed in that study were given for use in this one. Because the net's mesh size was quite large and also targeted plankton larger than copepods, the samples were filtered through a 0.5 mm sieve before use in this study. They were fixed in 2% paraformaldehyde in filtered seawater at 4 °C overnight. They were then washed twice in miliQ water, twice in 30% ethanol and

twice in 50% ethanol. After washing, they were transferred to 50% ethanol and stored at 4 °C until further processing.

Only a small number of samples were used in this study. Region boundaries were determined as described in section 2.2.1 and shown in Figure 2.1. Figure 2.4A shows the AMT22 cruise track with stations used for Illumina sequencing (red circles) and the station at which the clone library sample (white circle) was collected.

2.3.2 AMT24

2.3.2.1 *Pleuromamma* copepod sample collection

Collection of mesozooplankton samples for this study was done in collaboration with another group working on zooplankton ecology on board the AMT24 cruise, led by Erica Goetze (University of Hawaii).

Sampling was conducted at the pre-dawn stations using a double 0.71m-diameter bongo net with a 200μm mesh size (Figure 2.2A). Collection of samples was done at night to ensure sampling of the migratory community. A total of 45 oblique tows from 401-70m depth (with an average maximum of 305m depth) and surface were done using the bongo net system. The samples were collected for research on three distinct projects related to zooplankton ecology, undertaken by different teams participating in the AMT24 cruise (see AMT24 Cruise Report for more detail). Broadly, the themes were (1) Evolutionary ecology, (2) Copepod-associated bacterial communities (this study) and (3) Nauplii diversity, distribution and community structure.

This study focused on the collection of samples from the genus *Pleuromamma*, in particular on the two smaller species *P. gracilis/piseki* (PLGP) and *P. borealis* (PLBO). The genus *Pleuromamma* was selected because it is an important component of the pelagic zooplankton community. All of the species are strong vertical migrators and as such have an important role in nutrient export and marine biogeochemical cycles (Steinberg *et al*., 2000). The smaller species dominate the 0.5 to 1.5 mm size fraction of zooplankton samples (Landry *et al*., 2001). Stations with samples collected for this study are shown in Figure 2.4B and Table 2.2.

Immediately after the nets were brought on-board, they were washed with seawater to ensure most of the material was in the cod-ends, which were then emptied into buckets and stored on ice until further processing. Material from one net was immediately

preserved for molecular studies, while material from the second net was quantitatively split and half further processed for metagenomic and metagenetic studies. This research is being undertaken by researchers from the Goetze lab (University of Hawaii).

Samples used in this study came from the second half of the quantitative split of net 2 material. As soon as possible after collection, animals were individually identified and sorted by species. They were then fixed for three main purposes:

(1) Molecular analysis: next-generation sequencing and Fluorescence *in situ* hybridisation (FISH) (**Chapters 5 and 6**)
(2) Elemental analysis and dry weight (**Chapter 3 and 4**)
(3) Electron microscopy (**Chapter 3 and 6**)

For molecular analysis, individuals were grouped for fixation (between 5 and 30 animals) and fixed in 2% freshly prepared paraformaldehyde in sterile filtered seawater, either for 6h at room temperature or overnight at 4 °C. Then followed two one-hour washing steps in filtered seawater and 1h in 30% Ethanol. The samples were finally transferred into 50% Ethanol and stored at 4 °C until further processing.

Individuals for elemental and stable isotope analysis were also grouped prior to fixation and blotted dry, then transferred to cryotubes and immediately frozen at -80 °C until analysis.

Animals from selected stations, where enough material was available, were preserved for electron microscopy. After sorting, they were transferred into cold 2.5% glutaraldehyde and stored at 4 °C until further processing.

Table 2.2 AMT24 zooplankton stations sampled for this study. NT – Northern Temperate, NAG – North Atlantic Gyre, EQ – Equatorial Region, SAG – South Atlantic Gyre and ST – Southern Temperate Region.

Cruise-Station	Latitude	Longitude	Date Sampled	Region
ZPT_JR303_001	46° 23.033 N	10° 58.528 W	28/09/2014	NT
ZPT_JR303_002	44° 5.455 N	14° 54.735 W	29/09/2014	NT
ZPT_JR303_003	41° 46.292 N	18° 45.431 W	30/09/2014	NT
ZPT_JR303_004	39° 24.417 N	22° 28.231 W	01/10/2014	NT
ZPT_JR303_005	34° 44.92 N	26° 37.879 W	03/10/2014	NT
ZPT_JR303_006A	31° 52.296 N	27° 33.151 W	03/10/2014	NAG
ZPT_JR303_006	31° 17.711 N	27° 42.255 W	04/10/2014	NAG
ZPT_JR303_007	27° 2.978 N	28° 52.317 W	05/10/2014	NAG
ZPT_JR303_008A	24° 37.307 N	29° 44.774 W	05/10/2014	NAG
ZPT_JR303_008	24° 3.628 N	29° 52.967 W	06/10/2014	NAG
ZPT_JR303_009	20° 27.033 N	29° 15.365 W	07/10/2014	NAG
ZPT_JR303_010A	18° 19.88 N	28° 48.465 W	07/10/2014	NAG
ZPT_JR303_011	14° 12.596 N	27° 55.385 W	09/10/2014	EQ
ZPT_JR303_012A	11° 22.186 N	27° 19.932 W	09/10/2014	EQ
ZPT_JR303_013	7° 17.2674 N	26° 29.4594 W	11/10/2014	EQ
ZPT_JR303_014A	4° 24.617 N	25° 54.209 W	11/10/2014	EQ
ZPT_JR303_014	3° 48.2118 N	25° 46.638 W	12/10/2014	EQ
ZPT_JR303_015	0° 4.82 N	25° 52.62 W	13/10/2014	EQ
ZPT_JR303_016	3° 53.2188 S	25° 1.28 W	14/10/2014	EQ
ZPT_JR303_017	7° 28.2627 S	25° 2.2117 W	15/10/2014	EQ
ZPT_JR303_018A	10° 26.23 S	25° 2.836 W	16/10/2014	SAG
ZPT_JR303_018	11° 2.3303 S	25° 2.716 W	16/10/2014	SAG
ZPT_JR303_019	14° 39.566 S	25° 4.413 W	17/10/2014	SAG
ZPT_JR303_020A	17° 41.41 S	25° 5.47 W	18/10/2014	SAG
ZPT_JR303_020	18° 18.986 S	25° 4.979 W	18/10/2014	SAG
ZPT_JR303_021	20° 51.314 S	25° 4.633 W	20/10/2014	SAG
ZPT_JR303_022A	23° 51.4716 S	25° 2.7042 W	20/10/2014	SAG
ZPT_JR303_022	24° 27.443 S	25° 2.5063 W	21/10/2014	SAG
ZPT_JR303_023A	27° 8.0134 S	25° 0.0557 W	21/10/2014	SAG
ZPT_JR303_023	27° 45.66 S	25° 0.625 W	22/10/2014	SAG
ZPT_JR303_024A	30° 35.49 S	25° 48.288 W	22/10/2014	SAG
ZPT_JR303_025A	34° 11.0843 S	27° 12.4312 W	23/10/2014	ST
ZPT_JR303_026A	37° 11.533 S	28° 28.98 W	24/10/2014	ST
ZPT_JR303_028	41° 28.645 S	33° 51.465 W	27/10/2014	ST
ZPT_JR303_029	43° 0.797 S	37° 8.391 W	28/10/2014	ST

2.3.2.2 Microplankton sample collection

In order to assess the diversity of the *Pleuromamma* copepods' potential food source, samples of the microplankton size fraction were collected at the mid-day sampling stations (Figure 2.3B and Table 2.3). These were used to measure the trophic baseline isotopic values and provide insight into the community composition of the microplankton along the transect.

An *in-situ* size fractioning microplankton net (Micronet) with a double net frame was deployed to 200 m depth (Figure 2.2B). The vertical net hauls were deployed only when weather conditions were favourable (wind speed less than Beafourt force 6). The size fractions were >180 µm, 100-180 µm and 20-100 µm with samples from the latter two collected in acid washed bottles.

Immediately after collection, the two samples from each size fraction (i.e. 20-100 µm and 100-180 µm) were imaged live with a FlowCam (Fluid Imaging Technologies, Inc.). Further samples of each size fraction were then filtered onto GF/F filters using a low-pressure vacuum pump and frozen at -80° C until further analysis. For most stations, both net samples per size fraction could be filtered onto a single filter; however, in some cases in which the biomass in the sample was too dense, each net sample was filtered onto a single GF/F filter.

Table 2.3 AMT24 Micronet stations used in this study. NT – Northern Temperate, NAG – North Atlantic Gyre, EQ – Equatorial Region, SAG – South Atlantic Gyre and ST – Southern Temperate Region.

Cruise-Station	Latitude	Longitude	Date sampled	Region
MPV-JR303-007	33° 21.233 N	27° 5.075 W	03/10/2014	NAG
MPV-JR303-008	29° 53.92 N	28° 9.913 W	04/10/2014	NAG
MPV-JR303-009	26° 18.311 N	29° 18.311 W	05/10/2014	NAG
MPV-JR303-010	22° 39.529 N	29° 51.985 W	06/10/2014	NAG
MPV-JR303-012	16° 26.164 N	28° 24.290 W	08/10/2014	EQ
MPV-JR303-013	12° 50.641 N	27° 38.827 W	09/10/2014	EQ
MPV-JR303-014	9° 22.101 N	26° 55.259 W	10/10/2014	EQ
MPV-JR303-015	5° 53.948 N	26° 12.278 W	11/10/2014	EQ
MPV-JR303-016	2° 22.549 N	25° 29.174 W	12/10/2014	EQ
MPV-JR303-017	5° 19.34 S	25° 1.575 W	14/10/2014	EQ
MPV-JR303-018	8° 54.671 S	25° 2.560 W	15/10/2014	EQ
MPV-JR303-019	12° 29.686 S	25° 3.918 W	16/10/2014	SAG
MPV-JR303-020	16° 7.892 S	25° 5.680 W	17/10/2014	SAG
MPV-JR303-021	22° 18.985 S	25° 3.612 W	20/10/2014	SAG
MPV-JR303-022	25° 35.684 S	25° 2.145 W	21/10/2014	SAG
MPV-JR303-023	29° 9.236 S	25° 16.536 W	22/10/2014	SAG
MPV-JR303-024	32° 42.993 S	26° 37.912 W	23/10/2014	ST
MPV-JR303-025	40° 41.126 S	32° 6.196 W	26/10/2014	ST

Chapter 3

Molecular characterisation and morphological description of *Pleuromamma* copepods

3.1 Introduction

The calanoid copepod genus *Pleuromamma* (Metridinidae), with 11 described species, is an important component of the pelagic mesozooplankton community (Goswami *et al*., 1992; Beaugrand *et al.*, 2002). All species are pronounced vertical migrators and, as such, have an important role in nutrient and carbon export from the surface to deeper mesopelagic waters (Morales, 1999; Steinberg *et al*., 2000). The genus has a worldwide distribution (Halbert *et al.*, 2013; Goetze, 2011; Hirai *et al.*, 2015) and the species are abundant members of the migratory mesozooplankton community composition (Wood-Walker *et al.*, 2002; Gaard *et al.*, 2008). For example in the Pacific Ocean, *Pleuromamma* copepods represented 76% of all animals sampled in the euphotic zone at night (Al-Mutairi *et al.*, 2001).

The species found in the temperate and tropical Atlantic Ocean are *P. xiphias, P. abdominalis, P. gracilis, P. piseki and P. borealis*. Although they are all strong vertical migrators, their depth distribution depends on their size and life stage. Adults of all species, as well as copepodites, spend the night-time feeding at a depth of 0-100 m and migrate out of the euphotic zone during daytime to avoid predation (Ambler and Miller, 1987). The largest species, *P. xiphias* and *P. abdominalis*, can migrate up to 1500 m depth, while the smaller species *P. gracilis, P. piseki and P. borealis* inhabit a depth up to 500 m (Buskey *et al*., 1989). The smaller species have been found to dominate the 0.5 to 1.5 mm size fraction of zooplankton samples (Landry *et al*., 2001). Another aspect making this genus a suitable choice for this study is their easy identification in a bulk zooplankton sample, because of the distinctive black pigment spot on the left or right side of the second thoracic segment found on animals of both sexes (Blades-Eckelbarger and Youngbluth, 1988).

The focus of this study was on the three small-bodied species within the *Pleuromamma* genus. These are *P. gracilis* and *P. piseki* occuring predominantly in subtropical gyres, although they have also been observed in the equatorial convergence zone (Beaugrand *et al.*, 2002) and *P. borealis,* which is found in the tropical equatorial waters (Steur, 1932). All

three species are thought to be omnivorous, feeding on a range of phytoplankton, microzooplankton and detritus (Kleppel, 1993; Wilson and Steinberg, 2010; Teuber *et al.*, 2014). The identification of the three species is based on very slight morphological characteristics, based primarily on the shape of the genital pore (Steur, 1932; Ferrari, 1984). The identification of males and juveniles is difficult and this study focused mainly on female samples. Although the genus is considered systematically well-resolved, recent advances in molecular markers have shown that many cryptic species complexes exist in the genus, for example the *P. gracilis* and *P. piseki* complex (Halbert *et al.*, 2013). In the most recent study of the phylogenetic relationship between these two species, Halbert *et al.* (2013) found that there are in fact several molecular lineages co-occurring worldwide, with little ecological niche separation and little morphological differentiation.

The aim of this chapter is to characterise the phylogeny on a molecular level and describe the morphological characteristic of the studied *Pleuromamma* copepods. The phylogenetic relationship of samples used in this thesis is shown using the mitochondrial COI marker gene. Samples were imaged using electron microscopy in order to identify morphological characteristics of the feeding appendages that would help identify potential food sources and the inner morphology, in particular the gut morphology, is described in order to support the characterisation of the gut bacterial community in Chapter 6. Additionally, a small set of samples was analysed to determine the body size and condition of the *Pleuromamma* copepods.

3.2 Methods

3.2.1 Sample selection

Samples used in this part of the study came primarily from the AMT24 sampling effort (described in Chapter 2), although some COI sequences came from the AMT22 cruise. Analysis was done at the Max-Planck Institute for Marine Microbiology, Bremen.

The data shown here falls into three categories:

i) COI sequences for phylogenetic identification of individuals used in Chapters 5 and 6, and the copepod condition section of this chapter.
ii) Body size measurements of *P. gracilis/piseki* and *P. borealis* individuals.
iii) Electron microscopy of the outer and inner morphology of *P. gracilis* specimens.

3.2.2 COI amplification and phylogenetic analysis

An antenna was taken from all samples used in this study, and used for molecular identification of *Pleuromamma* individuals to complement the morphological identification of species. The mitochondrial COI gene was used for molecular identification using versatile primers L1384 (5' – GGT CAT GTA ATC ATA AAG ATA TTG G – 3') and H2612 (5' – AGG CCT AGG AAA TGT ATM GGG AAA – 3') developed by Machida *et al.* (2004). Other marker genes, e.g. mitochondrial COII (Halbert *et al.*, 2013), nuclear ITS (Goetze, 2003) or conserved 18S rDNA (Bucklin *et al.*, 2003; Goetze, 2003), could have been used, however, the COI gene was chosen because it had the most publically available data linked to *Pleuromamma* species (NCBI database) at the time of this project.

The PCR amplification reactions were carried out in a final volume of 50 µl using the Phusion Polymerase PCR Kit (NEB, USA) and following the manufacturer's recommendations. The reaction mix contained: 1x GC PCR buffer, 0.2 mM of each dNTP, 0.5 µM of each primer. An antenna, taken from copepods, was added to the PCR master mix without polymerase and the mixture heated at 95 °C for 5 min. Finally, 0.02 U/µl Phusion DNA polymerase was added to the mix. The temperature profile of the reaction was a 30 s initial denaturation step at 98 °C followed by 35 cycles of 10 s denaturation at 98 °C, 30 s of annealing at 42 °C and 45 s elongation at 72 °C; the PCR was completed with a final 10 min extension step at 72 °C. The PCR products were resolved on 1.4% agarose-TAE gels with a low mass DNA ladder (Invitrogen, Carlsbad, USA) and run for approximately 35min at 100 V. Agarose gels were post-stained for 20-30 min using Ethidium Bromide DNA stain and visualised under an UV light source. Amplified products of the correct length were purified using a QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) in preparation for Sanger sequencing. Purified samples were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit on the Genetic Analyzer Abiprism 3130 (Applied Biosystems, Foster City, USA).

Phylogenetic analysis of the sequences was done using Geneious v.8.1.7 software (Kearse *et al.*, 2012). The FindModel web analyser (Posada and Crandall, 1998) was used to select the best nucleotide substitution model. The model fitting the COI alignment best, based on the Akaike information criterion, was the General Time Reversible (GTR) model plus Gamma. Trees were also checked with different treeing algorithms (e.g. PHYML and Geneious tree builder) and produced comparable tree topologies.

Raw sequence reads were trimmed to remove primers and low quality or short reads were removed from further analysis. Several reference *P. gracilis*, *P. piseki* and *P. borealis* COI gene sequences were downloaded from the NCBI database (see accession numbers in tree). Sequences were aligned using MAFFT alignment (Katoh and Standley, 2013). Trees were built using Bayesian phylogenetic reconstruction with MrBayes v.3.2.2 (Huelsenbeck and Ronquist, 2001). A GTR model with Gamma-distributed rates of evolutions was used with 1,100,000 generations using four parallel Monte Carlo Markov chains and tree subsampling every 200 generations. A *Metridia lucens* (Metridinidae) COI gene sequence downloaded from NCBI was used as an outgroup.

3.2.3 Copepod condition measurements and elemental composition

Individual copepods photographed prior to stable isotope analysis were used for measurements of prosome length (PL) and calculation of the condition factor index (CFI). Five to seven individuals were pooled for elemental and stable isotope analysis (discussed in Chapter 4). Pooled animals were placed in pre-weighed tin cups and dried overnight at 70 °C. Tin cups were then weighed again to determine dry weight (DW) of pooled samples.

The PL of individual copepods was measured using ImageJ software. Average PL values were calculated for each pool of animals and the CFI was calculated according to the following equation (Ikeda *et al.*, 2006): $CFI = 1000 \times DW/(PL)^3$. This was done for pooled samples, because the information for DW was not available for individual copepods. Variables were plotted as boxplots using R statistical software version 3.2.1. More detailed statistical analysis was not performed due to variable and, in some cases, too small (EQ samples) sample sizes for reliable comparisons.

3.2.4 Electron microscopy and histology

Electron microscopy was done by Nikolaus Leisch (Symbiosis Department, MPI-MM) at the MPI-MM, Bremen.

Specimens were postfixed in 1% osmiumtetroxide and dehydrated in a graded ethanol series. Scanning electron microscopy specimens were critical point dried using a CPD 300 (Leica), mounted on stubs and gold coated with a ACE200 (Leica). For transmission electron microscopy, the samples were then transferred to 100% dry acetone and infiltrated using centrifugation (modified from McDonald, 2014) in 2ml tubes sequentially with 25%, 50%, 75% and 2x 100% Low-Viscosity resin (Agar Scientific). Samples were

placed into a pre-filled tube and centrifuged for 30 seconds with a bench top centrifuge (Heathrow Scientific, USA) at 2000g for each step. Finally, they were transferred into embedding moulds and polymerized at 60°C for 24h. Semi-thin (1µm) and ultra-thin (70 nm) sections were cut with an Ultracut UC7 (Leica Microsystem, Austria). Ultra-thin sections were mounted on formvar coated slot grids and contrasted with 0.5 % aqueous uranyl acetate (Science Services, Germany) for 20 min and 2 % Reynold's lead citrate for 6 min.

Samples were imaged at 10-20kV on a Quanta FEG 250 scanning electron microscope (FEI Company, USA) equipped with a STEM detector using xT microscope control software ver. 6.2.6.3123.

3.2.5 Supplementary information – Appendix A

Additional images of *Pleuromamma* histology supporting gut morphology reconstruction are shown in Appendix A, Figures A.1 and A.2. These include paraffin sections of *P. gracilis/piseki* and *P. borealis* and toluidine blue stained semi-thin sections of the TEM-prepared material.

3.3 Results

The *Pleuromamma* species studied were of similar size and ecology. They differ in distribution and have some morphological differences. Figure 3.1 shows examples of both species used in this study. *P. gracilis/piseki*, with a rounded genital pore (GP) is shown in Figure 3.1A, while *P. borealis,* with a skull-shaped GP, is shown in Figure 3.1B. The prosome length (PL) and urosome (UL) are also shown.

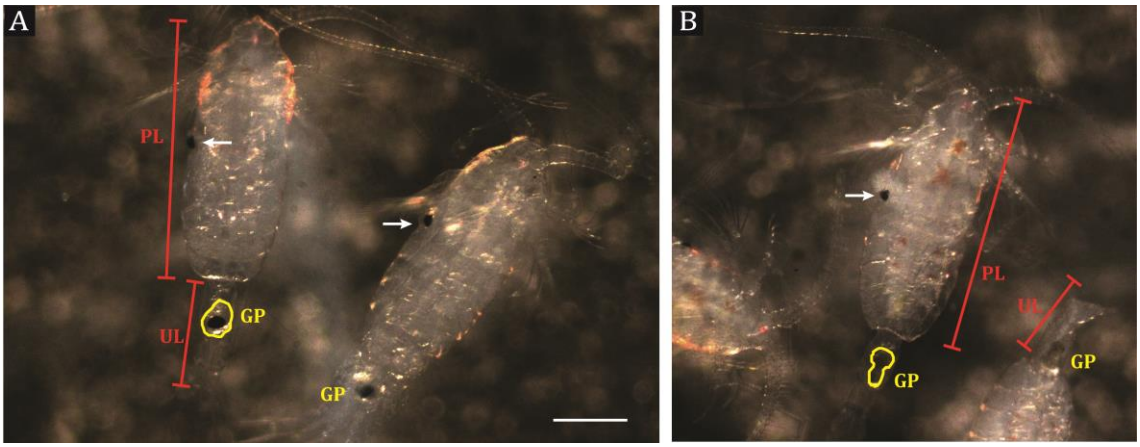
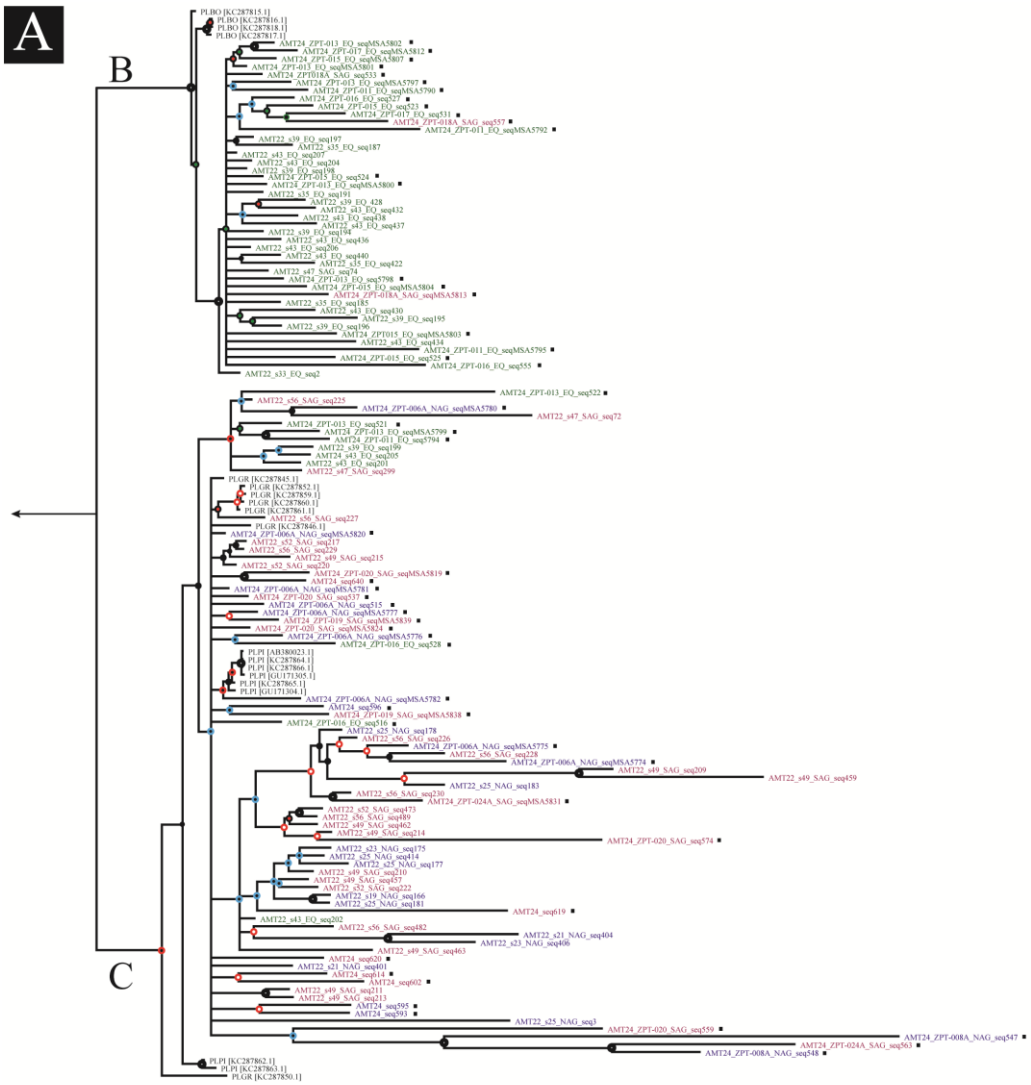


Figure 3.1 Example images of female *P. gracilis/piseki* (**A**) and *P. borealis* (**B**). Yellow outlines highlight difference in genital pore (GP) shape between the two species. White arrows show the pigment spot characteristic of the genus *Pleuromamma*. Red lines show prosome (PL) and urosome (UL) length. Scale: 0.5 mm.

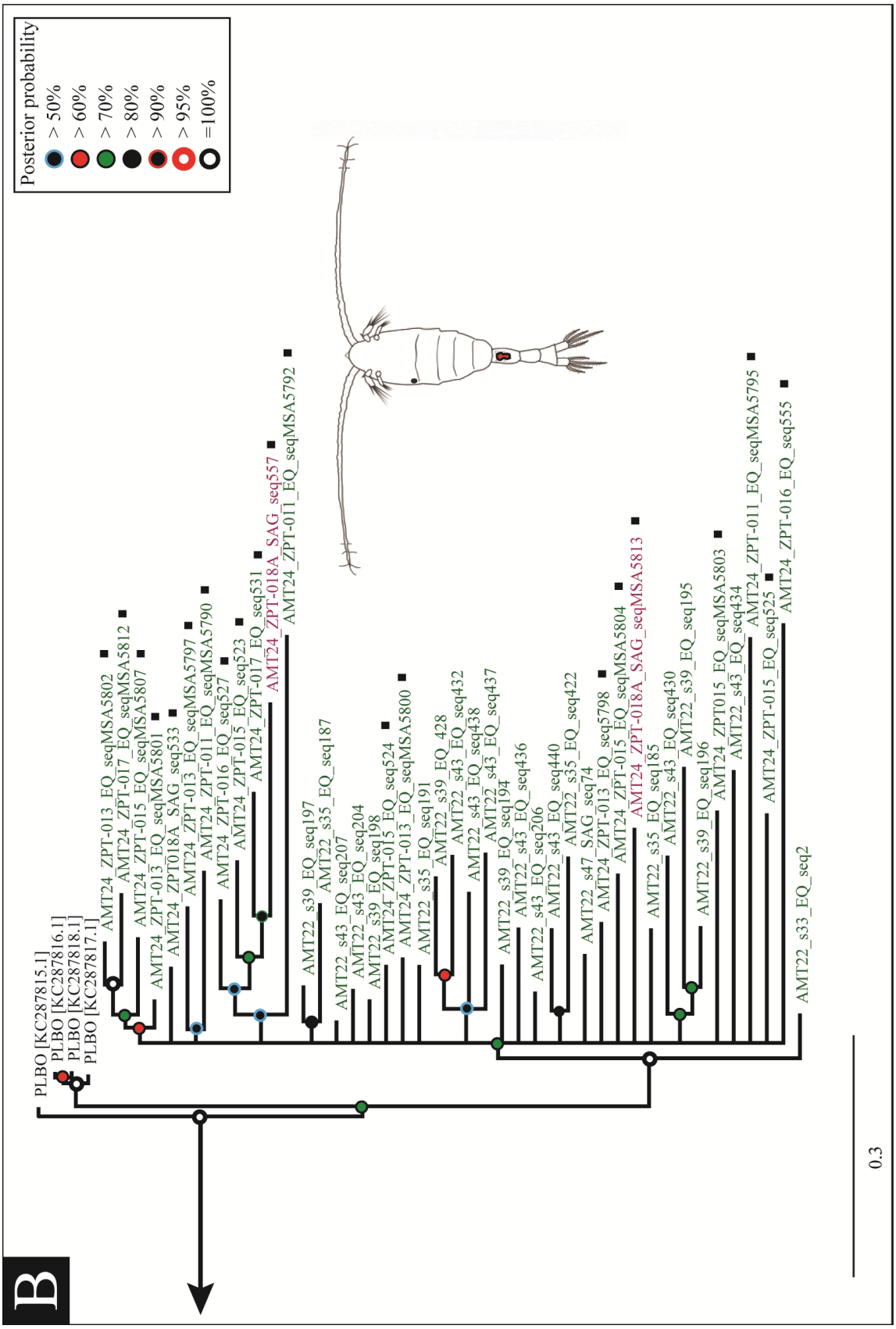
3.3.1 Phylogenetic relationship of *P. gracilis*, *P. piseki* and *P. borealis*

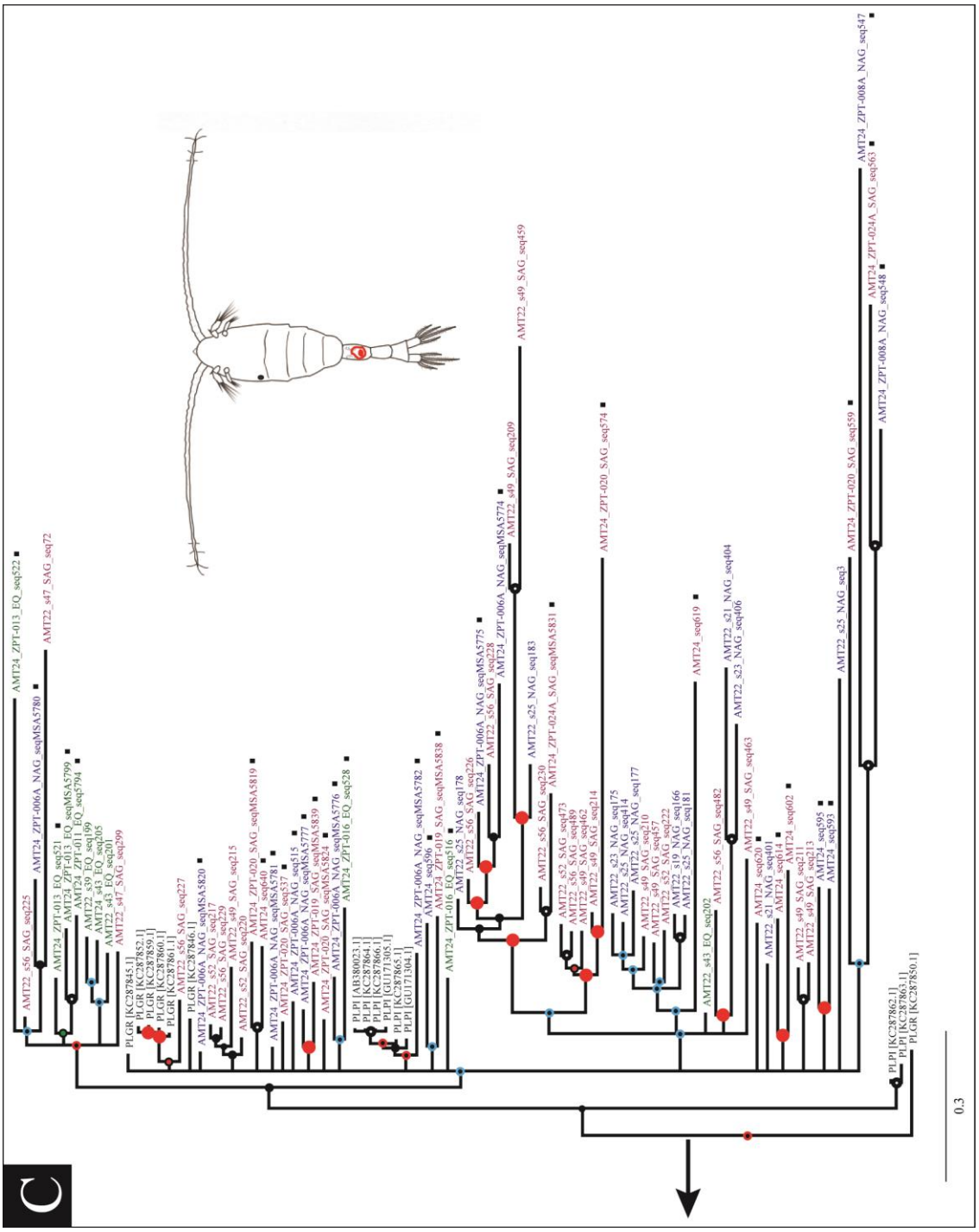
The phylogenetic analysis of mitochondrial COI genes amplified from copepod samples used in this study revealed two distinct groups. One was related to *P. borealis* sequences (PLBO), which consisted mostly of samples from EQ and two samples from a station in the SAG (AMT24_ZPT-018A) bordering the EQ region (Figure 3.2A). The second group (PLGP) consisted of a mix of sequences related to either *P. gracilis* or *P. piseki* (Figure 3.2B). The phylogeny within this group could not be resolved. The placement of the NCBI reference sequences in the tree agreed with BLAST results obtained for individual amplified COI sequences.



(continued on next page)

Figure 3.2 Bayesian inference tree of mitochondrial COI gene sequences showing the phylogenetic relationship among AMT22 and AMT24 samples used in this study. **A:** Overview tree. Details shown in separate panels. Two groups could be resolved: **B: PLBO** – *P. borealis* related sequences and **C (next page): PLGP** – *P. gracilis* and *P. piseki* related sequences. Regions are shown in colour (Blue – NAG, Green – EQ, Maroon – SAG). Posterior probability is shown at branch nodes and a *Metridia lucens* COI sequence was used to root the tree.


B
Posterior probability
> 50%
> 60%
> 70%
> 80%
> 90%
> 95%
=100%
PLBO [KC287815.1]
PLBO [KC287816.1]
PLBO [KC287818.1]
PLBO [KC287817.1]
AMT24_ZPT-013_EQ_seqMSA5802
AMT24_ZPT-017_EQ_seqMSA5812
AMT24_ZPT-015_EQ_seqMSA5807
AMT24_ZPT-013_EQ_seqMSA5801
AMT24_ZPT018A_SAG_seq533
AMT24_ZPT-013_EQ_seqMSA5797
AMT24_ZPT-011_EQ_seqMSA5790
AMT24_ZPT-016_EQ_seq527
AMT24_ZPT-015_EQ_seq523
AMT24_ZPT-017_EQ_seq531
AMT24_ZPT-018A_SAG_seq557
AMT24_ZPT-011_EQ_seqMSA5792
AMT22_s39_EQ_seq197
AMT22_s35_EQ_seq187
AMT22_s43_EQ_seq207
AMT22_s43_EQ_seq204
AMT22_s39_EQ_seq198
AMT24_ZPT-015_EQ_seq524
AMT24_ZPT-013_EQ_seqMSA5800
AMT22_s35_EQ_seq191
AMT22_s39_EQ_428
AMT22_s43_EQ_seq432
AMT22_s43_EQ_seq438
AMT22_s43_EQ_seq437
AMT22_s39_EQ_seq194
AMT22_s43_EQ_seq436
AMT22_s43_EQ_seq206
AMT22_s43_EQ_seq440
AMT22_s35_EQ_seq422
AMT22_s47_SAG_seq74
AMT24_ZPT-013_EQ_seq5798
AMT24_ZPT-015_EQ_seqMSA5804
AMT24_ZPT-018A_SAG_seqMSA5813
AMT22_s35_EQ_seq185
AMT22_s43_EQ_seq430
AMT22_s39_EQ_seq195
AMT22_s39_EQ_seq196
AMT24_ZPT015_EQ_seqMSA5803
AMT22_s43_EQ_seq434
AMT24_ZPT-011_EQ_seqMSA5795
AMT24_ZPT-015_EQ_seq525
AMT24_ZPT-016_EQ_seq555
AMT22_s33_EQ_seq2
0.3


C
AMT24_ZPT-013_EQ_seq522
AMT22_s56_SAG_seq225
AMT24_ZPT-006A_NAG_seqMSA5780
AMT22_s47_SAG_seq72
AMT24_ZPT-013_EQ_seq521
AMT24_ZPT-011_EQ_seqMSA5799
AMT24_ZPT-011_EQ_seq5794
AMT22_s39_EQ_seq199
AMT24_s43_EQ_seq205
AMT22_s43_EQ_seq201
AMT22_s47_SAG_seq299
PLGR [KC287845.1]
PLGR [KC287852.1]
PLGR [KC287859.1]
PLGR [KC287860.1]
PLGR [KC287861.1]
AMT22_s56_SAG_seq227
PLGR [KC287846.1]
AMT24_ZPT-006A_NAG_seqMSA5820
AMT22_s56_SAG_seq217
AMT22_s56_SAG_seq229
AMT22_s49_SAG_seq215
AMT22_s52_SAG_seq220
AMT24_ZPT-020_SAG_seqMSA5819
AMT24_seq640
AMT24_ZPT-006A_NAG_seqMSA5781
AMT24_ZPT-020_SAG_seq537
AMT24_ZPT-006A_NAG_seq515
AMT24_ZPT-006A_NAG_seqMSA5777
AMT24_ZPT-019_SAG_seqMSA5839
AMT24_ZPT-020_SAG_seqMSA5824
AMT24_ZPT-006A_NAG_seqMSA5776
AMT24_ZPT-016_EQ_seq528
PLPI [AB380023.1]
PLPI [KC287864.1]
PLPI [KC287866.1]
PLPI [GU171305.1]
PLPI [KC287865.1]
PLPI [GU171304.1]
AMT24_ZPT-006A_NAG_seqMSA5782
AMT24_seq596
AMT24_ZPT-019_SAG_seqMSA5838
AMT24_ZPT-016_EQ_seq516
AMT22_s25_NAG_seq178
AMT22_s56_SAG_seq226
AMT24_ZPT-006A_NAG_seqMSA5775
AMT22_s56_SAG_seq228
AMT24_ZPT-006A_NAG_seqMSA5774
AMT22_s49_SAG_seq209
AMT22_s49_SAG_seq459
AMT22_s25_NAG_seq183
AMT22_s56_SAG_seq230
AMT24_ZPT-024A_SAG_seqMSA5831
AMT22_s52_SAG_seq473
AMT22_s56_SAG_seq489
AMT22_s49_SAG_seq462
AMT22_s49_SAG_seq214
AMT24_ZPT-020_SAG_seq574
AMT22_s23_NAG_seq175
AMT22_s25_NAG_seq414
AMT22_s25_NAG_seq177
AMT22_s49_SAG_seq210
AMT22_s49_SAG_seq457
AMT22_s52_SAG_seq222
AMT22_s19_NAG_seq166
AMT22_s25_NAG_seq181
AMT24_seq619
AMT22_s43_EQ_seq202
AMT22_s56_SAG_seq482
AMT22_s21_NAG_seq404
AMT22_s23_NAG_seq406
AMT24_seq620
AMT22_s49_SAG_seq463
AMT22_s21_NAG_seq401
AMT24_seq614
AMT24_seq602
AMT22_s49_SAG_seq211
AMT22_s49_SAG_seq213
AMT24_seq595
AMT24_seq593
AMT22_s25_NAG_seq3
AMT24_ZPT-020_SAG_seq559
AMT24_ZPT-008A_NAG_seq547
AMT24_ZPT-024A_SAG_seq563
AMT24_ZPT-008A_NAG_seq548
PLPI [KC287862.1]
PLPI [KC287863.1]
PLGR [KC287850.1]
0.3

3.3.2 Body size and condition of animals along the AMT24 cruise track

Measurements of prosome length (PL) made on individual copepods from different regions along the AMT24 cruise track showed sizes ranging from 1 mm to 1.6 mm (Figure 3.3). The variability in body size was higher in the gyre animals, i.e within the *P. gracilis/piseki* species complex. They ranged in size from 0.9-1.6 mm, with an average size of 1.3 mm (± 0.11 mm). *P. borealis* animals ranged in size from 1.2-1.4 mm, with an average 1.3 mm (± 0.06 mm).The species could not be determined in some cases, such as male specimens or when the shape of the genital pore was not clear. The condition (CFI) of animals was very similar between the gyres and EQ (Table 3.1).

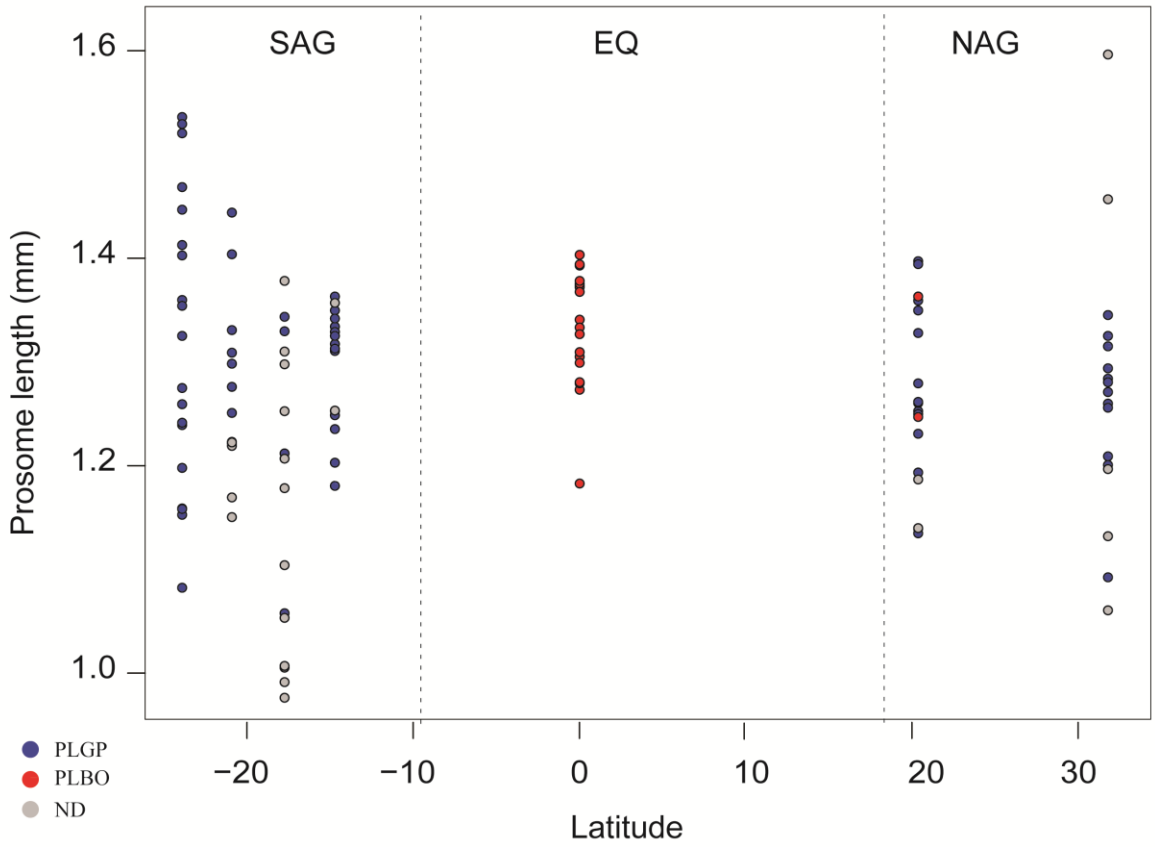


Figure 3.3 Body size of *P. gracilis/piseki* (PLGP) and *P. borealis* (PLBO) copepods from selected stations along the AMT24 cruise track. Regions: North Atlantic Gyre – NAG, Equatorial Region – EQ, South Atlantic Gyre – SAG.

Table 3.1 Mean and variance of prosome length (PL) and condition factor index (CFI) by region. N – number of measurements. ind – number of individuals used in CFI calculation. North Atlantic Gyre – NAG, Equatorial Region – EQ, South Atlantic Gyre – SAG.

Variable	Region	N (ind)	Mean (± SD)	Variance
PL (mm)	NAG	35	1.27 ± 0.11	0.01
	EQ	20	1.34 ± 0.06	0.00
	SAG	66	1.27 ± 0.13	0.02
CFI	NAG	6 (30)	19.34 ± 5.73	32.87
	EQ	2 (14)	18.59 ± 1.31	1.73
	SAG	12 (51)	21.42 ± 9.26	85.81

3.3.3 Morphological description

The outer morphology of a female *P. gracilis* was investigated using SEM (Figure 3.4). Of particular interest were the feeding appendages (FA). A set of outer FA adapted for concentrating food towards the mouth could be seen in the images (Figure 3.4B). The size of the filaments may be appropriate for capturing particles or marine snow. The appendages in the mouth looked adapted for crushing particles, with interlocking teeth seen deep inside the mouth (Figure 3.4C and D).

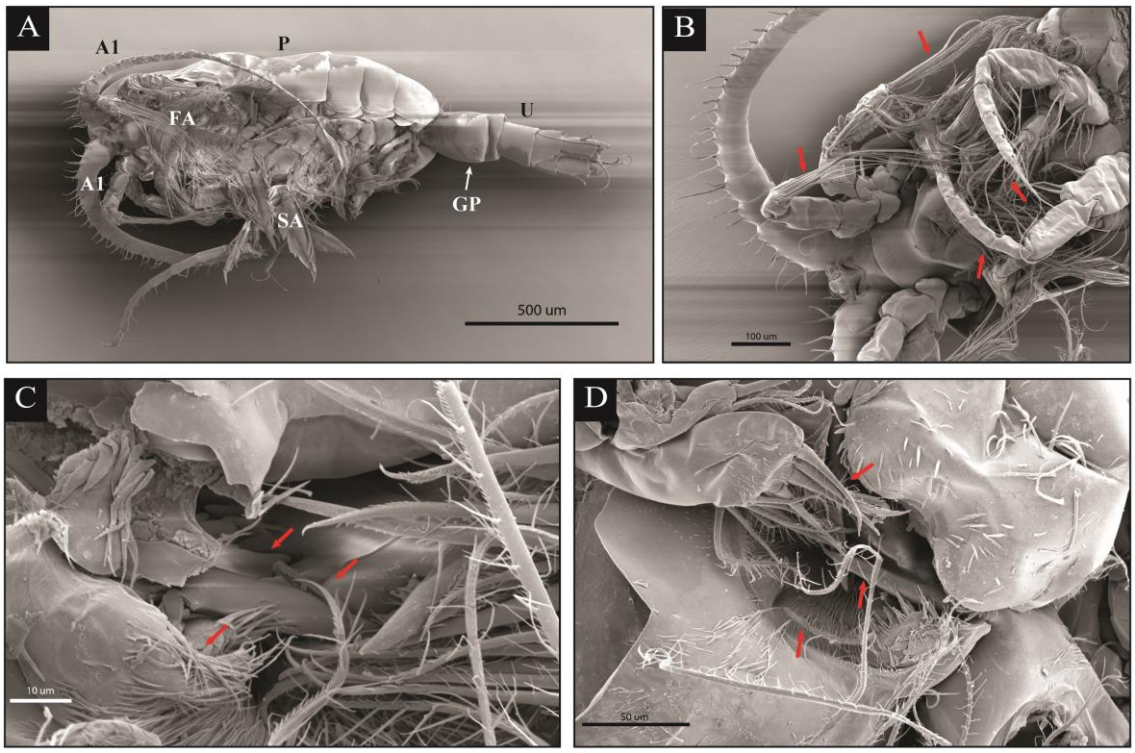
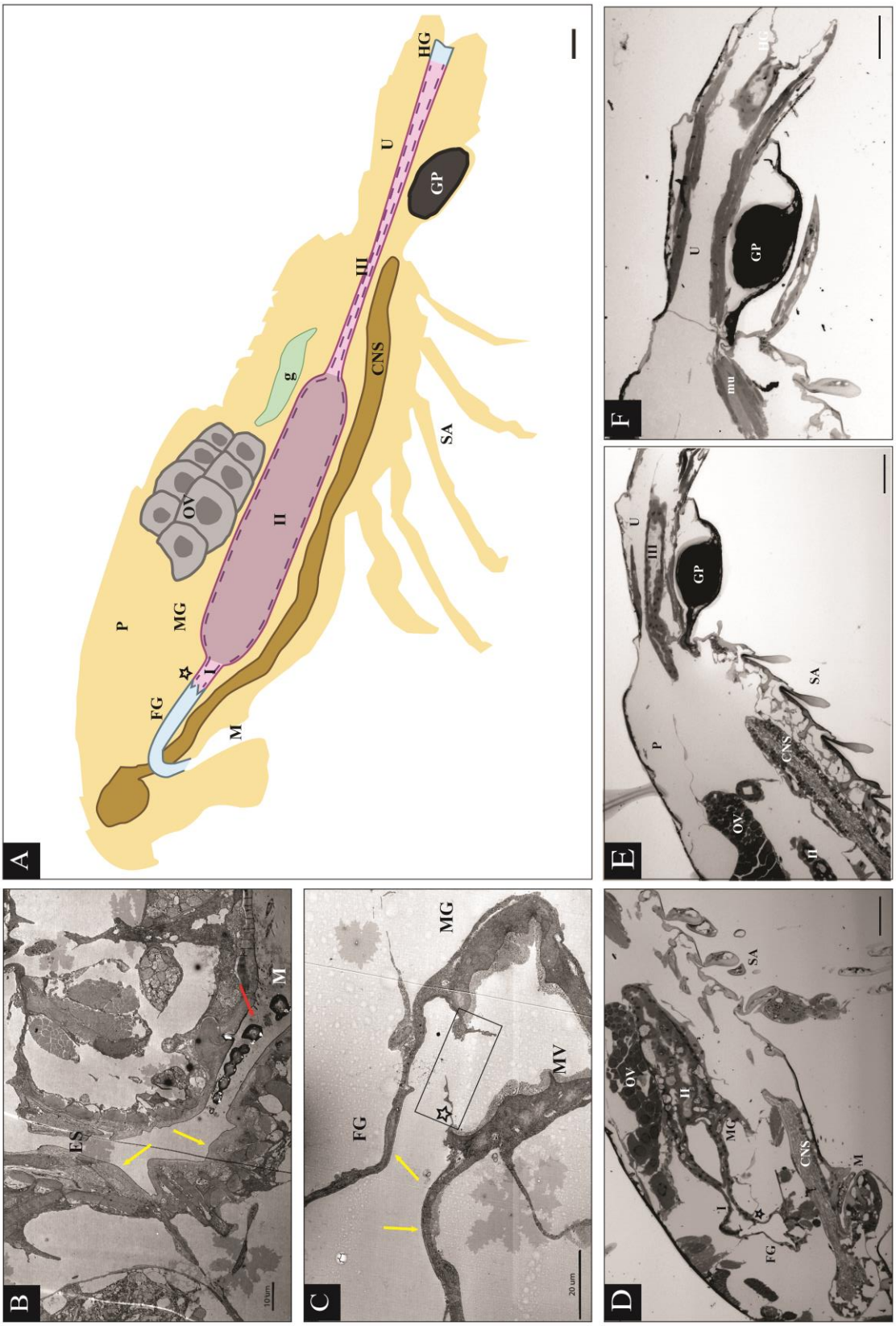


Figure 3.4 Outer morphology of a *P. gracilis* female copepod **(A)** and close up images of outer **(B**) and inner **(C and D**) feeding appendages **(red arrows).** A1 – first antennae. FA – feeding appendages. SA – swimming appendages. GP – genital pore. P – prosome. U – urosome.

The inner morphology of a female *P. gracilis* individual was determined using a combination of microscopy approaches and histological staining. The main aim was to determine the gut morphology for use in further gut microbe analysis. However, other features were also identified and are shown in Figure 3.5.

The morphology of the gut could be divided into distinct regions (Figure 3.5). The foregut was characterised by the presence of a cuticle lining of the inner gut wall (Figure 3.5B, C). This is expected to be the case for the hindgut as well, but the hindgut could not be identified with certainty in any of the examined sections. This suggested that it is highly reduced (Figure 3.5A and F). The transition between the foregut and the midgut was clearly visible (Figure 3.5). The largest gut region was the midgut, which could be further divided into three areas. Areas I and III had a simple tube-like appearance and a single cell epithelial layer (Figure 3.5 D and E). The most distinct was Area II, with large cells and invaginations of the tissue lining, enlarging the surface area (Figure 3.5D). The epithelium of the entire midgut was covered by microvilli. Other features such as the ventral central nervous system (CNS), ovaries (OV) and extensive musculature, could also be observed in the histological sections of the female *P. gracilis* sample.

Figure 3.5 (next page) Inner morphology of a *P. gracilis* female copepod. **A:** Schematic showing details of gut morphology and other features observed in histological sections. Scale: 50 μm. **B:** Close up of mouth. **C**: close up of transition between FG and MG. **D:** Overview of anterior prosome. **E:** Overview of posterior prosome and urosome. **F:** Overview of urosome with HG**. Abbreviations** FG – foregut. MG – midgut. HG – hindgut. Blue – cuticle lined tissue in FG and HG. Purple – three regions of MG (I-III). Darker purple – area II with tissue invaginations. Dashed line – tissue with microvilli. Star – separation between FG and MG. ES – esophagus. Red arrows – inner feeding appendages. Yellow arrows – cuticle lining. SA – swimming appendages. GP – genital pore. P – prosome. U – urosome. OV – ovaries. CNS – central nervous system. M – mouth. mu – muscle tissue.


A
HG
U
GP
III
CNS
g
SA
OV
II
P
MG
FG
I
M
B
ES
M
10um
C
MG
FG
MV
20 um
D
OV
II
SA
MG
I
FG
CNS
M
E
U
III
GP
SA
P
CNS
OV
II
F
HG
U
GP
III

3.4 Discussion

The molecular identification and morphological characteristics of *Pleuromamma gracilis/piseki* and *P. borealis* copepods, collected primarily along the AMT24 cruise transect, were presented in this chapter. The molecular identification and phylogenetic relationship among the samples was explored using partial mitochondrial COI (mtCOI) gene sequences. All samples used in this study were identified based on morphological characteristics (Figure 4.1) and this was complemented with molecular identification. Two clear groups of sequences were identified in the phylogenetic tree (Figure 3.2). One was composed of samples mainly from the EQ region and they were related to *P. borealis* reference sequences. This is consistent with previous observations of the distribution of this species (Steur, 1932). The second group contained sequences related to both *P. gracilis* and *P. piseki* but more detailed separation of these two species was not possible based on data available from this study. It is likely that the phylogeny of the whole genus *Pleuromamma* is not as well resolved as previously thought and many studies have reported the presence of cryptic species complexes, including the *P. garcilis/piseki* group (Ferarri and Fornshell, 2010; Halbert *et al.*, 2013). The group will therefore be referred to as the *P. gracilis*/*piseki* (PLGP) and not treated as two separate species in further analyses.

The aim of this chapter was not to do an in-depth population genetics analysis of the small-bodied *Pleuromamma* species, but rather to complement the morphological identification of the copepods with molecular data. Therefore other molecular marker genes (such as the 18S rDNA gene, the nuclear ITS or mitochondrial COII gene) were not used to complement the data shown here. Nevertheless, any future studies exploring the data presented here further, should consider using other marker genes to get a more robust phylogenetic analysis. As discussed in the Methods section of this chapter, the COI gene marker was chosen because there is a lot of data in public databases already available. It was also considered appropriate baceuse the sequence is very stable across taxa (Avise *at al.*, 1987; Feral, 2002) and universal primers (e.g. the ones developed by Machida *et al.*, 2005) could be used without having to devote time to developing specific primer sequences. However, certain limitations are associated with the use of mitochondrial genes as markers that apply in particular to population genetic studies. In crustaceans, the occurance of NUMTs (nuclear mitochondrial pseudogenes) is common and can affect DNA barcoding studies by overestimating unique sequences (Song *et al.*, 2008). This was not an issue here, because the mtCOI sequences were only used to complement morphological identification. An issue that did occur was repeated poor

quality of certain sequenced samples and this could be linked to the amplification of duplicate sequences caused by the presence of NUMTs.

The body size and condition of a small set of copepods were measured for stations from three regions along the AMT24 cruise track. The body size of measured animals varied more within the gyre regions than in the EQ region. An explanation for this could be that samples from the gyres belonged to the *P. gracilis/piseki* species complex (Halber *et al.,* 2013) and it is therefore likely that the different species within the complex vary in body size. Additionally, more samples were measured from gyre station and also more male specimens were included in the selection. The difference between male and female animals is also a likely source of variation in the measured prosome lengths. Due to the small number of males used in the study, further comparisons and statistical analysis was not done. The Condition Factor Index (CFI), as proposed by Ikeda *et al*. (2006), was used to calculated values showing the condition of the animals. The obtained values were within the range of values reported for other mesopelagic and *Pleuromamma* copepods (Ikeda *et al*., 2006). However, the sample number was very different between species and regions, therefore a comparison was challenging due to high variance. Other trophic biomarker have been used to assess the condition of copepods, for example the presence of storage lipids (Teuber *et al.*, 2014). This approach could be challengeing for active migratory copepods, such as *Pleuromamma*, that are characterised by high protein and low lipid content (Morris and Hopkins, 1983). The ratio of elemental carbon and nitrogen (C/N) would be a more appropriate measure of condition for these animals, with low C/N values indicating a high protein, i.e. mussle, content (Morris and Hopkins, 1983).

Finally, the outer and inner morphology of *P. gracilis* was investigated using a range of microscopy and histology approaches. The images of outer feeding appendages showed two different mechanisms for optimising food intake. The appendages around the outside of the mouth were long, filamentous and full of setae, which suggests the ability to detect, capture and concentrate food toward the mouth. Teeth seen inside the mouth, on the other hand, suggest the capacity to crush and break up particles and phytoplankton cells, for example diatom frustules. These teeth-like structures are typical of copepods feeding on phytoplankton (Michels and Gorb, 2015). *Pleuromamma* copepods in the N Pacific Gyre have been found to feed on a variety of food from autotrophic picoplankton, most likely ingested as part of aggregates, and microzooplankton (Wilson and Steinberg, 2010). These groups are a likely food source for *Pleuromamma* copepods in the Atlantic (Chapter 2 and 4). The trophic position of other *Pleuromamma* species was assessed in the Atlantic Ocean

using lipid biomarkers and stable isotope analysis and the study showed that *Pleuromamma* are omnivores (Teuber *et al.*, 2014).

There have been very few recent studies of the morphology and functioning of copepod guts. However, the structure of the gut in calanoid copepods is quite uniform, with three main regions: chitinised foregut and hindgut, and midgut with a microvilli-covered epithelial layer (Hallberg and Hirche, 1980). Although, the hindgut was not found in the TEM images in this study, both the chitinised foregut and midgut with microvilli were observed in the *P. gracilis* copepod (Figure 3.5B and C). The midgut can then be further divided into three parts, with the middle section composed of large vacuolated cells (Arnoud *et al.*, 1978). These cells are thought to have a role in intracellular digestion and are accompanied by cells for digestive enzyme secretion; while cells in the elongated lower part of the midgut have a resorptive function (Hassett and Blades-Eckelbarger, 1995). These characteristics were seen in several histological sections in this study (Figure 3.5C and Figure A.1 in Appendix A).

Chapter 4

Microplankton and *Pleuromamma* copepod elemental and stable isotope analysis along AMT24

4.1 Introduction

The presence of gut microbial communities has been documented in many different animal species. These communities have been shown to play an important role in the host's digestion by synthesising molecules the host cannot, and maintaining general health (McFall-Ngei *et al*., 2013). The main factors influencing the community composition of a gut microbiome are the host's diet, environment and gut morphology (Engel and Moran, 2013). Understanding the trophic ecology of the host can thus help to explain community structure of the gut microbiota and their geographical distribution patterns.

Oceanic copepods exhibit a large range of feeding behaviours. Many species are active predators feeding on other mesozooplankton and larger prey (e.g. fish larvae). Others are generalists and can feed on both motile food and non-motile particles (Kleppel, 1993; Laakmann and Auel, 2010). Food webs in oligotrophic waters typically have a high proportion of omnivores and opportunistic predators, since primary production and biomass are limited due to low nutrient supplies. Important food sources for mesozooplankton are microplankton species in the size range between 20 - 200 µm, such as heterotrophic flagellates, ciliates and diatoms (Stoecker and Capuzzo, 1990; Calbet and Siaz, 2005). Determining the feeding behaviour and dietary composition of copepods can be assessed using laboratory feeding experiments, microscopy and FlowCam measurements of prey before and after incubation (Ide *et al.*, 2008). However, this can be challenging for open-ocean pelagic copepods, primarily because of the difficulty of routinely sampling remote areas and maintaining of laboratory cultures of organisms of interest (Njestgaard *et al.*, 2003). Molecular approaches such as the use of trophic markers can therefore be employed to indirectly determine the feeding ecology of pivotal members of the pelagic trophic web.

In recent years, stable isotopes have become widely used tools for studying many aspects of trophic ecology (e.g. dietary habits and trophic relationships) and animal movement (e.g. migration or foraging strategies) (reviewed by Boecklen *et al*., 2011). The use of trophic markers, such as the ratio of stable isotopes of carbon ($\delta^{13}C$) or nitrogen

($\delta^{15}N$), relies on the premise that there is differential retention of the heavy and light isotopes, depending on the metabolic pathways or trophic position of the organism (Michener and Lajtha, 2007).

The usefulness of stable isotopes lies in their predictable incorporation into predators' tissues from the diet. Isotope values in tissues of predators therefore vary from those of their prey in a step-wise manner and can provide valuable trophic information (DeNiro and Epstein, 1978; DeNiro and Epstein, 1981). Isotopic compositions of primary producers at the base of the food web can also vary spatially, and thus predator tissues can contain information of the geographical location of the animal during the period of tissue synthesis (Graham *et al.*, 2010). The $\delta^{15}N$ of consumers is typically enriched in ^{15}N and thus a higher $\delta^{15}N$ value is measured in upper trophic levels than at their source. This average trophic enrichment (TEF) has historically been considered to be relatively consistent with an increase by approximately 3.4 ‰ per trophic level (Minagawa and Wada, 1984). However, more recent studies have found that TEF can vary among systems and mostly depends on diet, species and body condition (Vander Zanden *et al*., 1997; Vanderklift and Ponsard, 2003). The $\delta^{15}N$ of primary producers can differ across spatial scales due to the use of different nutrient sources, with contrasting $\delta^{15}N$ values (Montoya *et al.*, 2002). Conversely, changes in $\delta^{13}C$ values between different levels in the trophic web are usually more stable and on the order of 0 to 1 ‰. As such, they are useful in tracing the carbon source of the food web (McCutchan *et al*., 2003, Caut *et al.*, 2009). Differences in phytoplankton $\delta^{13}C$ can be due to variation in community composition and cell geochemistry (e.g. cell geometry, mechanism of CO_2 uptake etc.) (Rau *et al.,* 1982; Popp *et al.*, 1998).

It is especially important to understand the carbon and nitrogen isotopic composition at the base of the food web when attempting to infer the trophic position or feeding patterns of consumers between different ecosystems. Choosing a suitable base is one of the biggest challenges in trophic isotope ecology (Post, 2002; Caut *et al.*, 2009). Primary producers often have quicker doubling times and therefore incorporate isotopes over a different timescale to their predators. This can create spatial and temporal discrepancies in the isotope signatures of both the prey base and consumers (Vanderklift and Ponsard, 2003; Landrum *et al*, 2011). In many cases, it is also difficult or impractical to determine the exact food source of a consumer. Such is the case with many zooplankton species that feed on a variety of microplankton and detrital material.

In this chapter, the isotopic and elemental compositions of microplankton in the 20 – 100 µm and 100 – 180 µm size fractions were investigated by analysing stable isotopes of carbon and nitrogen. These data were then compared to a set of *Pleuromamma* isotopic measurements. Additionally, the microplankton community composition of the two size fractions was determined using a FlowCAM approach. The aim was to examine the spatial profile of the microplankton community and use it as a reference dataset for assessing *P. gracilis/piseki* and *P. borealis* trophic ecology of samples collected on the AMT24 research cruise.

4.2 Materials and Methods

4.2.1 Sample collection

Microplankton samples for stable isotope and FlowCam analysis were collected at 18 stations along the AMT24 cruise track (Figure 4.1) for two different size fractions of microplankton: 20-100 µm ($Micronet_{20}$) and 100-180 µm ($Micronet_{100}$).

P. gracilis/piseki and *P. borealis* samples from two stations in the NAG, three station in EQ and two stations from SAG were used to determine the stable isotope composition along the AMT24 cruise track (Figure 4.1 and Table 4.1).

Details of sample collection can be found in Chapter 2. Further processing and analysis of samples was done at the Max-Planck Institute for Marine Microbiology (MPI-MM) in the Department for Microbial Ecology and Department for Biogeochemistry. Stable isotope measurements were done by Clara Martinez Perez and Julien Dekaezemacker (Department for Biogeochemistry, MPI-MM).

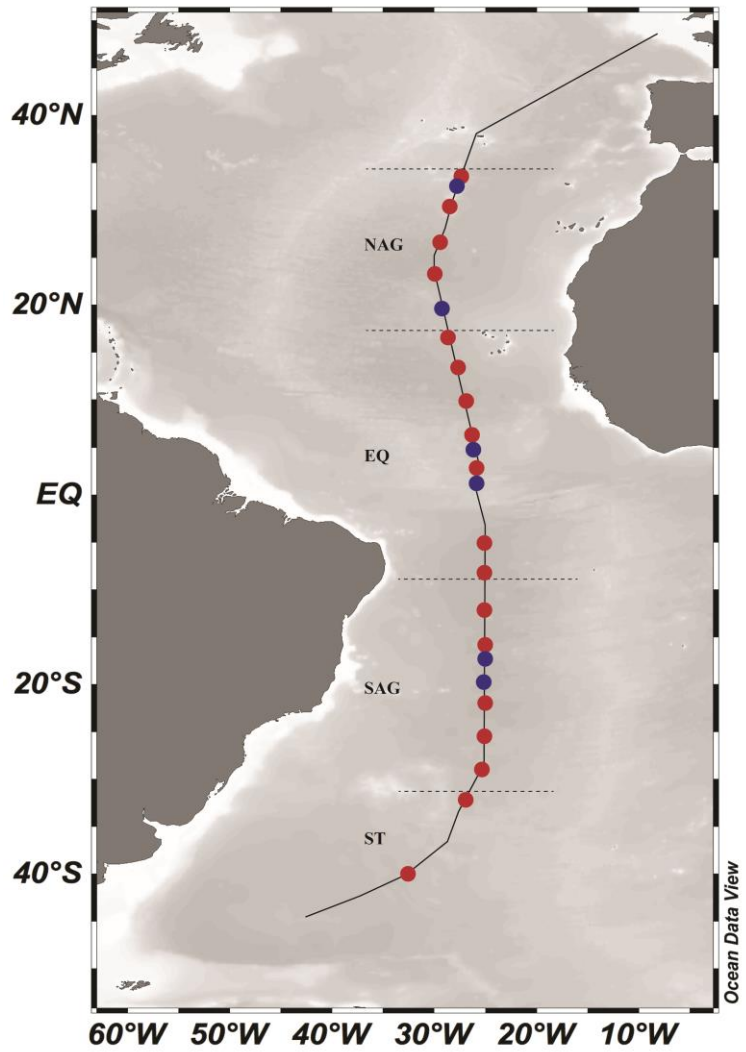


Figure 4.1 Samples used in Chapter 4. Micronet stations along the AMT24 cruise track are marked in red and copepod stations in blue. Dashed lines show region boundaries (North Atlantic Gyre – NAG, Equatorial Region – EQ, South Atlantic Gyre – SAG, Southern Temperate – ST).

Table 4.1 Stations used for stable isotope measurement of *P. gracilis/piseki* (PLGP) and *P. borealis* (PLBO) copepods. N – number of replicates per station. Ind – number of pooled individuals. North Atlantic Gyre – NAG, Equatorial Region – EQ, South Atlantic Gyre – SAG.

Station	Region	Latitude	Longitude	Copepod species	N (ind)
ZPT-006A	NAG	31° 52.296 N	27° 33.151 W	PLGP	3 (5-6)
ZPT-009	NAG	20° 27.033 N	29° 15.365 W	PLGP	3 (6)
ZPT-015	EQ	0° 4.82 N	25° 52.62 W	PLGP/PLBO	2 (7)
ZPT-016	EQ	3° 53.2188 S	25° 1.28 W	PLGP/PLBO	2 (6)
ZPT-020A	SAG	17° 41.41 S	25° 5.47 W	PLGP	3 (6)
ZPT-022A	SAG	23° 51.4716 S	25° 2.7042 W	PLGP	3 (6)

4.2.2 FlowCam image processing

Images obtained with the FlowCam (Fluid Technologies, Inc.; Sieracki *et al.*, 1998) were further analysed using Visual Spreadsheet software (Fluid Technologies, Inc.). Counts of different groups of microplankton were collected from FlowCam images. Groups were based on ease of identification and expected abundance in samples. Due to time constraints, only one net replicate and only stations used in stable isotope analysis (SIA) were analysed to determine microplankton community composition, so these are approximate estimates of community composition.

The numerical abundances per ml (Aml^{-1}) were calculated for each station per size fraction according to the following equation:

Aml^{-1} = (Particle count/ Total volume imaged [μm^3]) x ($1x10^{12}$ [μm^3]/ 1 [ml]).

Total volume imaged was calculated as "number of images taken during a single run" multiplied by the "volume per picture". Both of these values were determined from FlowCam software. The numerical abundance values were then normalised to the volume actually filtered through each net and each fraction at each station, taking into account concentrating of the sample. These abundances were then converted to numerical abundance per cubic meter (m^3).

4.2.3 Stable isotope and elemental C and N measurements

Micronet GF/F filters were acidified overnight over fuming HCL in a desiccator and then oven dried overnight at 70°C before being transferred to tin cups for analysis. Filters were divided into quarters prior to analysis because high material content could disrupt sample measurements.

Individual *Pleuromamma* copepods were photographed before five to seven individuals were pooled for isotope analysis. Pooled animals were placed in pre-weighed tin cups and dried overnight at 70 °C. Tin cups were then weighed again to determine dry weight (DW) of pooled samples.

Organic carbon and nitrogen content and bulk carbon and nitrogen isotopic compositions were measured using a Carbon-Hydrogen-Nitrogen (CHN) Elemental Analyzer (Thermo Flash EA 1112) coupled with a continuous flow isotope ratio mass spectrometer (Thermo Delta Plus XP, Thermo Fisher Scientific, Schwerte, Germany). Caffeine standards, calibrated against international standards, were measured every six

samples to ensure the accuracy of the measured values. Measurement errors were calculated for the entire set of standards over each mass spectrometer run for $\delta^{13}C$ and $\delta^{15}N$. Measurement runs, where the standard deviation of the caffeine standards exceeded 0.2 delta notations were treated with caution.

Isotope values were reported in the delta notation and calculated according to the following equation:

$$\delta^{13}C \text{ or } \delta^{15}N = ([R_{sample}/R_{standard}] - 1) \times 1000.$$

R is the $^{13}C/^{12}C$ or $^{15}N/^{14}N$ and the δ value shows the deviation of the measured ratio from an international standard in parts per thousand (‰). The standard for $\delta^{13}C$ is PDB limestone and atmospheric nitrogen is used as the standard for $\delta^{15}N$ (Newton, 2010).

Filters for Micronet samples were not combusted prior to sample collection. To control for measurement bias from the filter, a set of nine filters from the same box as those used for sample collection was treated and analysed in the same way as the samples described above. All control samples were unmeasurable for all parameters and it was concluded that the uncombusted filters would not bias Micronet sample measurements.

4.2.4 Data analysis

The integrated peak for carbon and nitrogen (provided by the Elemental Analyzer) of caffeine standards can be compared to the known mass of the standards analysed to create a factor (ugC/Area or ugN/Area). The average of all conversion factors in a measurement run was used to calculate the total carbon and nitrogen content per filter, according to the following equation:

$$\mu g(C \text{ or } N)_{1/4 \text{ filter}} = AreaAll(C \text{ or } N)_{1/4 \text{ filter}} \times \mu g(C \text{ or } N)/AreaAll.$$

$AreaAll(C \text{ or } N)_{1/4 \text{ filter}}$ is the integrated peak of the measured sample and µg(C or N)/AreaAll is calculated as an average value from the standards measured alongside the samples. Once values were calculated per filter quarter, they were added to obtain an organic carbon or nitrogen value per filter. An error occurred during the analysis of 8 stations (16 samples), which resulted in measurements of only three instead of four replicates per filter. In these cases, the mean value of the filter quarters was added to the sum of organic carbon or nitrogen to obtain a value per filter. The sum values were then divided by the volume of water filtered through the net: ~10,000 L when both net samples were filtered onto one filter or ~5,000 L when each net sample was filtered onto its own

filter. This gave the concentration of organic carbon and nitrogen per sample, expressed in μgL^{-1}. Measurements obtained from filter quarters were used to calculate the mean and standard deviation for the $\delta^{13}C$ and $\delta^{15}N$ per size fraction and station of Micronet samples.

Statistical analysis was done for four measured variables: $\delta^{13}C$, $\delta^{15}N$, μgCL^{-1} and μgNL^{-1}. Variables were tested for normality (q-q plots) and equal variance using Leven's tests. Firstly, a one-way ANOVA was used to test for significant differences between the two size fractions along the transect. The variation in trophic fractionation was calculated for both isotope ratios as the Trophic Enrichment Factor (TEF):

$$\Delta\delta^{15}N = \delta^{15}N_{100} - \delta^{15}N_{20} \text{ and } \Delta\delta^{13}C = \delta^{13}C_{100} - \delta^{13}C_{20}.$$

Secondly, a one-way factorial ANOVA was used to compare differences among biomes. Stations within the regions along the transect were grouped for analysis, and region (NAG, EQ, SAG, ST) was used as a factor. A Tukey's HSD test was performed for all significant results to determine the significant comparisons. All analyses were done in the R statistical software version 3.2.1.

The mean and standard deviation of copepod sample replicates was calculated for $\delta^{13}C$ and $\delta^{15}N$. Because there were a small number of copepod measurements no further statistical analysis was done.

The isotopic niche width of the two size fractions in different regions was also examined using the SIBER package (Jackson *et al.*, 2011) within the SIAR statistical package (Parnell *et al.*, 2010) in R software version 3.2.3. $\delta^{13}C$ and $\delta^{15}N$ values for all measured samples (filter quarter values instead of mean values per station) were plotted as bi-plots, representing the δ-space and isotopic niche space of groups of samples. Using this method it is possible to distinguish organisms based on their isotopic signature on both the trophic ($\delta^{15}N$) and primary carbon source ($\delta^{13}C$) level. This allows an estimation of trophic diversity and a comparison between the isotopic niches of different populations (Layman *et al.*, 2007; Jackson *et al.*, 2011). Samples were grouped according to region and size fraction; however, the EQ region was split into two due to the clear shift in $\delta^{15}N$ north and south of the Equator. The two stations from ST region were excluded from this analysis due to small sample size (Jackson *et al.* (2011) recommend a minimum of 10 samples). Eight groups were thus used in the analysis: NAG[20], NAG[100], N-EQ[20], N-EQ[100], S-EQ[20], S-EQ[100], SAG[20] and SAG[100]. Standard Ellipse Areas (SEAc: corrected for small sample sizes) and convex hulls were plotted for individual groups (SEAc included 40% of data, while convex hulls represented the total area). Bayesian

inference was then used to calculate and compare posterior distributions of ellipse areas (Jackson *et al.*, 2011).

4.3 Results

4.3.1 Microplankton community composition

The major groups present in both Micronet size fractions were Dinoflagellates (with *Ceratium* sp. dominating), copepod nauplii, adult copepods, *Trichodesmium* sp. (Cyanobacteria), Ciliates, Diatoms, Radiolaria, Detrital material (e.g. fecal pellets). Example low-resolution images of the groups obtained from the FlowCam software are shown in Figure 4.2. The estimated abundances of all groups differed among groups within regions but remained quite stable across regions (Figure 4.3). An exception was *Trichodesmium* sp., which showed a peak in abundance (57% contribution of all groups) in the EQ region.

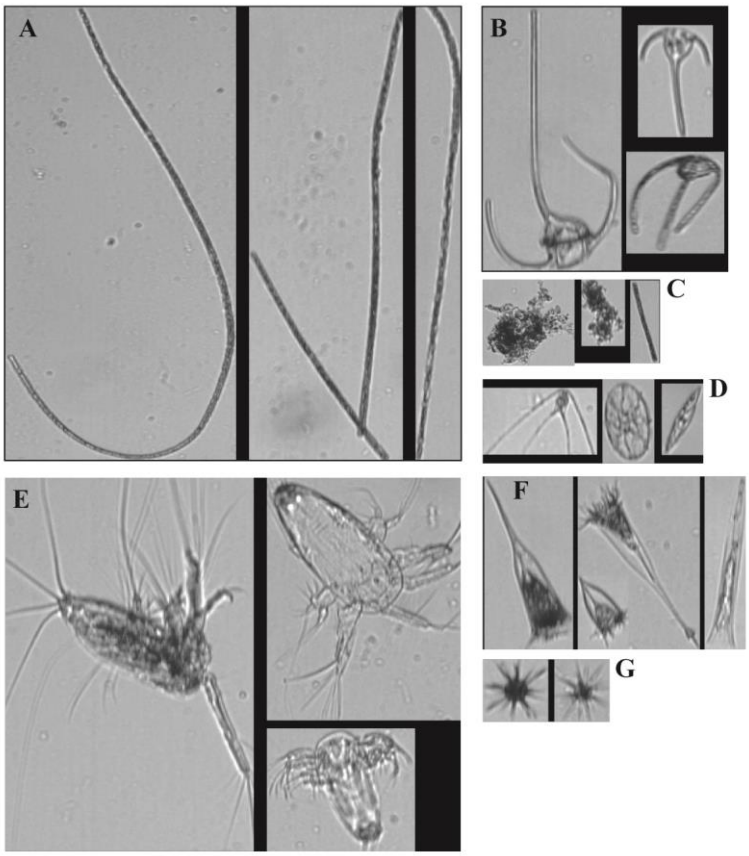


Figure 4.2 Example FlowCam images of microplankton groups identified in Micronet$_{20}$ samples. **A**: individual trichomes of the cyanobacterium *Trichodesmium* sp.; **B**: most dominant representative of the Dinoflagellates group, *Ceratium* sp.; **C**: example images of detrital material; **D**: different representatives of Diatoms; **E**: copepod nauplii; **F**: live Ciliates (Tintinnids); **G**: Radiolaria.

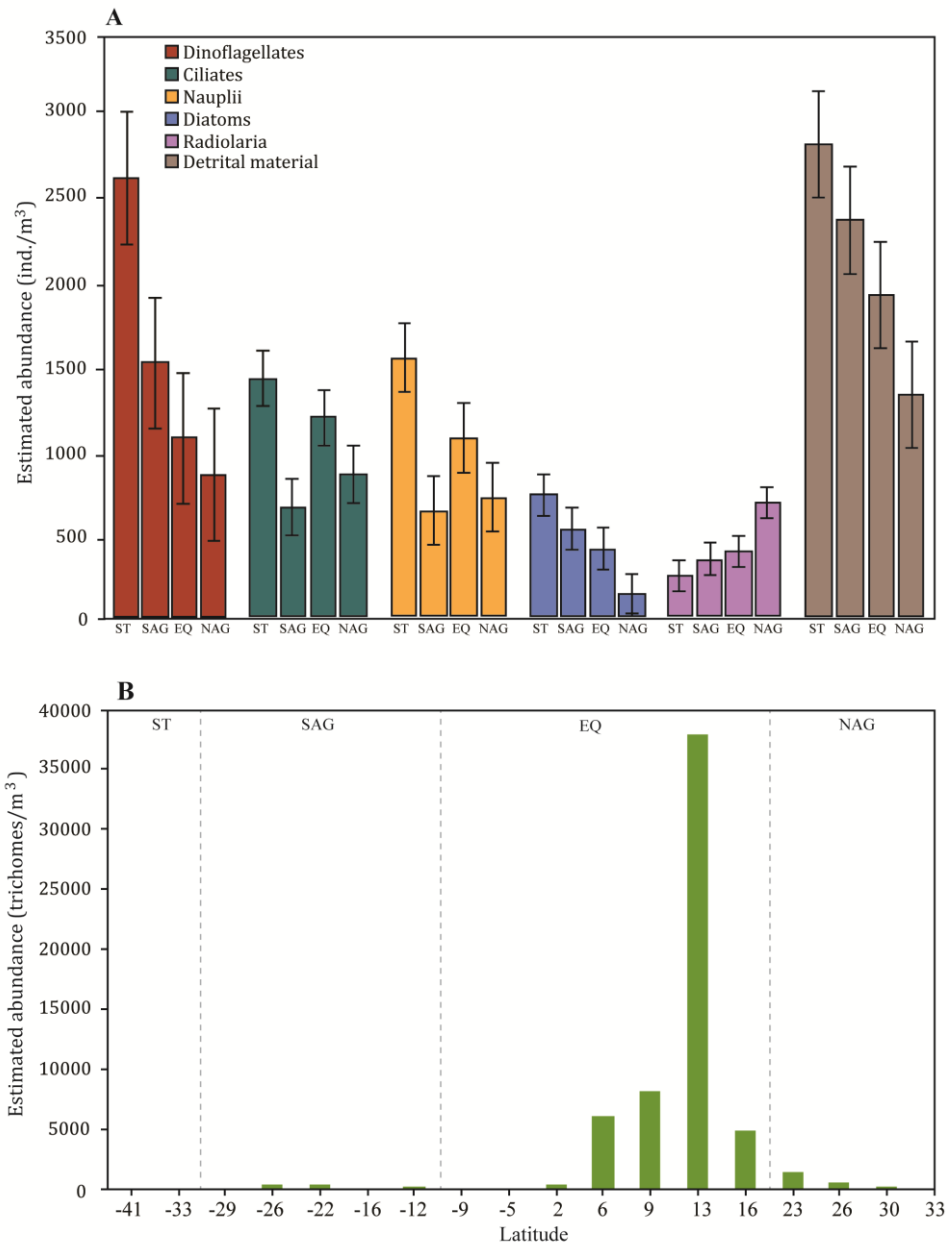


Figure 4.3 Estimated abundances of different groups identified in the Micronet$_{20}$ FlowCam samples. **A**: Abundance estimates for major groups by region. Error bars represent the Standard Error. **B**: *Trichodesmium* sp. estimated abundances along the transect. Latitude going from South to North, dashed lines represent region boundaries.

4.3.2 Microplankton carbon and nitrogen content

The carbon and nitrogen biomass along the transect had high biomass peaks in the EQ region and ST region, with lower values in both Gyre regions (Figure 4.4). On average, the biomass was higher in the smaller (Micronet$_{20}$) size fraction, except for the ST region, in which biomass was higher in Micronet$_{100}$ (Table 4.2). There was an unusually high value of both carbon and nitrogen recorded at station MPV-013 (~12° N) in Micronet$_{20}$ size

fraction. This can probably be attributed to a higher concentration of *Trichodesmium* sp. sampled at that station than at other stations. The comparison between the two size fractions (Table 4.3) did not differ significantly for either carbon or nitrogen biomass (one-way factorial ANOVA, $F_{1,34}$=3.465 with p=0.071 for carbon and $F_{1,34}$=0.7 with p=0.41 for nitrogen). Comparisons among regions (using both size fractions together) were significant for both elements (Table 4.3). In the case of carbon biomass ($F_{3,32}$=5.19, p<0.01) the only significant comparison was between the EQ region and SAG. Whereas for nitrogen biomass ($F_{3,32}$=4.53, p<0.001) comparisons between EQ and both Gyres, as well as between the SAG and the ST region, were significant.

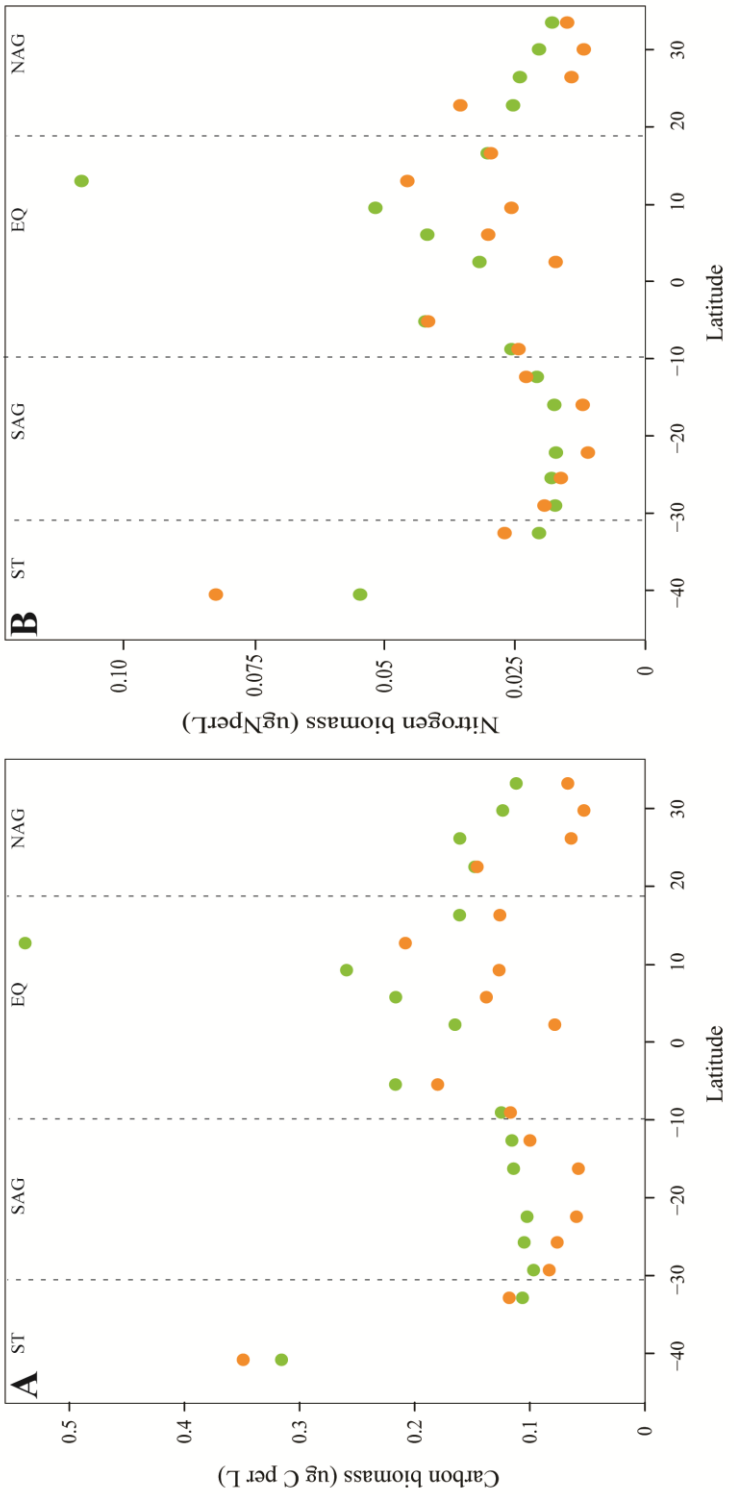


Figure 4.4 Variability of carbon and nitrogen content of Micronet samples along the latitudinal AMT24 transect (summed values across filter replicates and normalised for volume of water filtered). **A**: Carbon content in μgL^{-1} and **B**: Nitrogen content in μgL^{-1}. Green: $Micronet_{20}$. Orange: $Micronet_{100}$. Vertical dashed lines separate the different region (labels on top).

Table 4.2 Summary of carbon and nitrogen content (µgL^{-1}) showing: Mean (± SD), Variance, Leven's Test statistics for each size fraction across the AMT24 regions. The regions are: North Atlantic Gyre (NAG), Equatorial region (EQ), South Atlantic Gyre (SAG) and Southern Temperate (ST). Values in parentheses show the number of stations used in each region. Significant p values in bold.

Variable	Region (N)	Mean (± SD)	Variance	Leven's Test
20-100 um size fraction				
µgCL^{-1}	NAG (4)	0.136 ± 0.022	0	$F_{3,14}$=1.641, p=0.225
	EQ (7)	0.24 ± 0.139	0.019	
	SAG (5)	0.107 ± 0.008	0	
	ST (2)	0.211 ± 0.148	0.022	
µgNL^{-1}	NAG (4)	0.022 ± 0.003	0	$F_{3,14}$=1.492, p=0.26
	EQ (7)	0.047 ± 0.028	0	
	SAG (5)	0.018 ± 0.002	0	
	ST (2)	0.037 ± 0.024	0	
100-180 um size fraction				
µgCL^{-1}	NAG (4)	0.083 ± 0.044	0.002	$F_{3,14}$=6.821, **p<0.01**
	EQ (7)	0.139 ± 0.043	0.002	
	SAG (5)	0.075 ± 0.017	0	
	ST (2)	0.233 ± 0.163	0.027	
µgNL^{-1}	NAG (4)	0.019 ± 0.011	0	$F_{3,14}$=7.198, **p<0.01**
	EQ (7)	0.030 ± 0.01	0	
	SAG (5)	0.016 ± 0.005	0	
	ST (2)	0.054 ± 0.039	0.002	

Table 4.3 One-way factorial ANOVA with biomass parameters comparing both size fractions between each other and the combined dataset for each variable with with region as the explanatory factor. A Tukey's HSD post hoc test was done to determine significant comparisons among regions (North Atlantic Gyre – NAG, Equatorial region – EQ, South Atlantic Gyre – SAG and Southern Temperate – ST). Significant p values in bold.

Variable	ANOVA	Significant comparison
Comparison between size fractions		
µgCL^{-1}	$F_{1,34}$=3.465, p=0.071	n/a
µgNL^{-1}	$F_{1,34}$=0.7, p=0.41	n/a
Comparison among regions		
µgCL^{-1}	$F_{3,32}$=5.19, **p<0.01**	EQ-SAG
µgNL^{-1}	$F_{3,32}$=4.53, **p<0.001**	NAG-EQ, EQ-SAG, SAG-ST

4.3.3 Isotopic composition of microplankton and copepods

The $\delta^{15}N$ and $\delta^{13}C$ values measured in *P. gracilis/piseki* and *P. borealis* samples followed the Micronet baseline values (Figure 4.5). They were very similar to the Micronet values in the EQ region and the SAG, whereas they were slightly above the baseline in the NAG. Table 4.4 shows mean values and standard deviation of measured parameters averaged by region. Due to small samples size statistical comparison was not made and the following section describes the Micronet dataset in detail.

Table 4.4 Mean values (± SD) of dry weight (DW), $\delta^{13}C$, $\delta^{15}N$ and C/N ratio parameters of *Pleuromamma* spp. by region. The regions are: North Atlantic Gyre (NAG), Equatorial region (EQ), South Atlantic Gyre (SAG). Values in parentheses show number of replicates used in each region (each replicate consisted of a pool of 5-7 individuals).

Variable	Region (N)	Mean (± SD)
DW (mg)	NAG (6)	0.229 ± 0.065
	EQ (4)	0.278 ± 0.086
	SAG (6)	0.273 ± 0.113
$\delta^{13}C$	NAG (6)	-22.68 ± 1.73
	EQ (4)	-21.83 ± 1.86
	SAG (6)	-21.67 ± 1.61
$\delta^{15}N$	NAG (6)	3.75 ± 0.8
	EQ (4)	6.6 ± 0.5
	SAG (6)	3.75 ± 0.77
CN ratio	NAG (6)	5.4 ± 0.8
	EQ (4)	5.3 ± 0.3
	SAG (6)	4.2 ± 0.2

The largest variability along the transect was observed in $\delta^{15}N$ (Figure 4.5A and Table 4.5). However, Leven's test showed that populations within biomes had an equal variance ($F_{3,14}$=2.13 with p = 0.141 for $Micronet_{20}$ and $F_{3,14}$=1.44 with p = 0.274 for $Micronet_{100}$). The lowest values of $\delta^{15}N$ were recorded in the NAG with a mean value of 1.68‰ (± 0.57‰) in the small size fraction ($Micronet_{20}$) and 2.29‰ (± 0.13‰) in the larger size fraction ($Micronet_{100}$). The measurements ranged between 2.7‰ (± 0.2‰) at the northern edge of the region (approximately 33° N) to only 0.9‰ (± 0.7‰) at the border with the Equatorial region (approximately 22° N). The calculated diazotroph N contribution was highest for this region and contributed 45% (± 11.3%) in $Micronet_{20}$ and 42% (± 4.9%) in $Micronet_{100}$.

The EQ region had a clear transition in $\delta^{15}N$ from below 3‰ north of the Equator to between 5-7‰ south of it (for both size fractions). This can be seen in the high variance of 4.69‰ ($Micronet_{20}$) and 4.25‰ ($Micronet_{100}$) when the whole region is regarded as one. The contribution of diazotroph N in this region was 44% (± 6.5%) for $Micronet_{20}$ and 36.5% (± 2.8%) in $Micronet_{100}$. However, only values north of the Equator contributed to this regional average because the values from stations south of the Equator were negative and thus not included in the calculation.

The $\delta^{15}N$ isotope values in the SAG were lower than south of the Equator, but did not drop to below 3‰. The mean value in $Micronet_{20}$ was 3.77‰ (± 0.3‰) and 4.31‰ (± 0.7‰) in $Micronet_{100}$. Additionally, the percentage of N contributed through diazotrophy was much lower than the previous two regions, with values of 13.6% (± 7.8%) for $Micronet_{20}$ and 14.8% (± 8.9%) for $Micronet_{100}$. The ST region was again more variable, although only two stations were measured in this region, with very different $\delta^{15}N$ values. This was observed in the high variance of 2.33‰. In both size fractions, lower $\delta^{15}N$ values (3.3‰ in $Micronet_{20}$ and 4.2‰ in $Micronet_{100}$) were recorded in the southernmost station (MPV-025 at 40° S), while a value around 5.5‰ was observed in both size fractions for the station bordering the SAG. The diazotroph contribution was only calculated for the southernmost station and was 20.6 % for $Micronet_{20}$ and 17% for $Micronet_{100}$.

The carbon stable isotope ratios across all samples were much more variable, with high variability even within individual samples (Figure 4.5B and Table 4.5). Leven's test showed unequal variance among populations within regions for $Micronet_{20}$ ($F_{3,14}$=10.22, $p<0.01$) and equal variance for $Micronet_{100}$ ($F_{3,14}$=2.45, $p>0.1$). There was significant variation among regions ($F_{3,32}$=4.05, $p<0.01$), but the only significant differences were found between the EQ region and the SAG. For $Micronet_{20}$, average $\delta^{13}C$ values were lower in the EQ region (-21.93‰ ± 0.5‰) than in other regions, which had similar average $\delta^{13}C$ values from one to another (~ -22‰). For $Micronet_{100}$, average $\delta^{13}C$ values in the SAG (-23.43‰ ± 0.77‰) showed the greatest difference from others. The highest variability was recorded between the two ST stations.

Variation in the trophic fractionation of $\delta^{15}N$ and $\delta^{13}C$ between the two Micronet size fractions, calculated as the TEF (Δ), showed a very small increase in $\Delta^{15}N$ (average value across all samples 0.7‰ ± 0.5‰) with the highest difference calculated for the EQ region (0.9‰ ± 0.5‰). Similar calculations were made for $\Delta^{13}C$ (average across samples -0.34‰ ± 0.6‰).

The isotopic composition of $\delta^{15}N$ and $\delta^{13}C$ values between the two size fractions did not differ significantly (one-way factorial ANOVA, $F_{1,34} = 1.66$ with $p = 0.206$ for $\delta^{15}N$ and $F_{1,34} = 1.55$ with $p = 0.222$ for $\delta^{13}C$; Table 4.5). Both isotopic signatures showed significant differences among regions along the transect ($F_{3,32} = 4.53$, $F_{3,32} = 4.05$ with $p<0.01$ for $\delta^{15}N$ and $\delta^{13}C$, respectively). In the case of $\delta^{15}N$, significant differences were found between the NAG and the EQ region, between the two oligotrophic Gyres and between the NAG and the ST region. The only significant difference for $\delta^{13}C$ was between the EQ region and the SAG (Table 4.6).

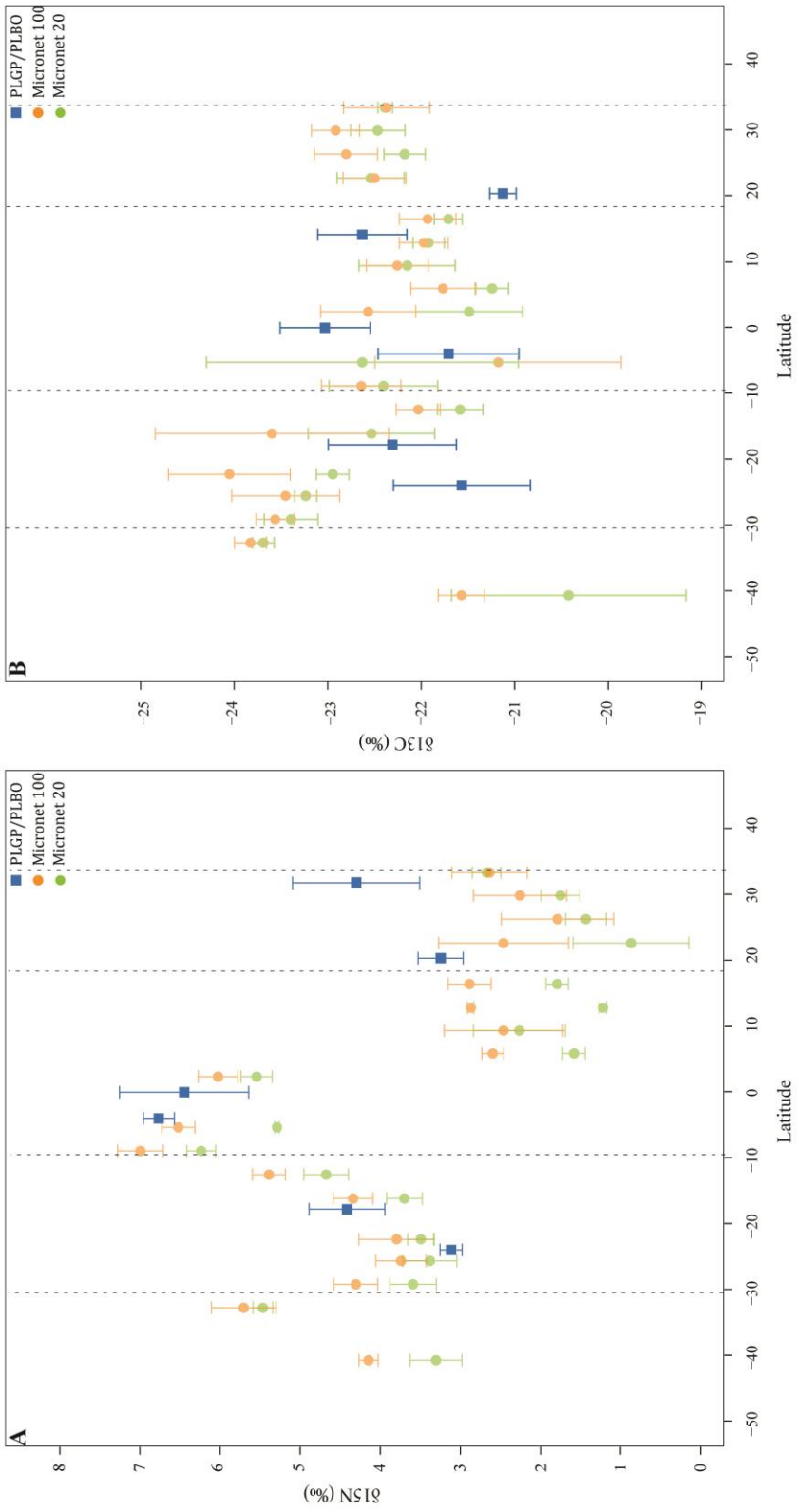


Figure 4.5 Variability of $\delta^{15}N$ and $\delta^{13}C$ along the AMT24 cruise track (averaged values across filter replicates). **A**: $\delta^{15}N$ isotopic signature and **B**: $\delta^{13}C$ isotopic signature. Green: $Micronet_{20}$. Red: $Micronet_{100}$. Blue: *Pleuromamma* copepods. Vertical dashed lines separate the different regions. The error bars represent the standard deviation of the filter quarters.

Table 4.5 Summary of $\delta^{15}N$ and $\delta^{13}C$ parameters showing: Mean (± SD), Variance and Leven's Test statistics for each size fraction across the AMT24 regions. The regions are: North Atlantic Gyre (NAG), Equatorial region (EQ), South Atlantic Gyre (SAG) and Southern Temperate (ST). Values in parentheses show number of stations used in each region. Significant p values in bold.

Variable	Region (N)	Mean (± SD)	Variance	Leven's Test
20-100 um size fraction				
$\delta^{15}N$	NAG (4)	1.68 ± 0.75	0.57	$F_{3,14}$=2.134, p=0.142
	EQ (7)	3.42 ± 2.16	4.69	
	SAG (5)	3.77 ± 0.52	0.27	
	ST (2)	4.39 ± 1.53	2.33	
$\delta^{13}C$	NAG (4)	-22.39 ± 0.16	0.02	$F_{3,14}$=10.22, **p<0.01**
	EQ (7)	-21.93 ± 0.5	0.25	
	SAG (5)	-22.74 ± 0.72	0.52	
	ST (2)	-22.06 ± 2.31	5.34	
100-180 um size fraction				
$\delta^{15}N$	NAG (4)	2.29 ± 0.37	0.13	$F_{3,14}$=1.44, p=0.274
	EQ (7)	4.34 ± 2.06	4.25	
	SAG (5)	4.31 ± 0.66	0.44	
	ST (2)	4.92 ± 1.1	1.21	
$\delta^{13}C$	NAG (4)	-22.65 ± 0.26	0.07	$F_{3,14}$=2.45, p=0.11
	EQ (7)	-22. 04 ± 0.5	0.25	
	SAG (5)	-23.43 ± 0.77	0.59	
	ST (2)	-22.7 ± 1.60	2.55	

* Contribution only from stations north of the Equator
** Only southernmost station in ST

Table 4.6 One-way factorial ANOVA with $\delta^{15}N$ and $\delta^{13}C$ parameters comparing both size fractions between each other and the combined dataset for each variable with with region as the explanatory factor. A Tukey's HSD post hoc test was done to determine significant comparisons among regions (North Atlantic Gyre – NAG, Equatorial region – EQ, South Atlantic Gyre – SAG and Southern Temperate – ST). Significant p values in bold.

Variable	ANOVA	Significant comparison
Comparison between size fractions		
$\delta^{15}N$	$F_{1,34}$=1.66, p=0.21	n/a
$\delta^{13}C$	$F_{1,34}$=1.55, p=0.22	n/a
Comparison among regions		
$\delta^{15}N$	$F_{3,32}$=4.53, **p<0.01**	NAG-EQ, NAG-SAG, NAG-ST
$\delta^{13}C$	$F_{3,32}$=4.05, **p<0.01**	EQ-SAG

4.3.4 Isotopic niche width

Comparison of microplankton populations among the 8 defined groups confirmed that there are differences among populations in different regions along the AMT24 cruise track (Figure 4.6). Groups containing both size fractions from NAG and N-EQ formed a cluster with smaller areas (SEAc), in comparison to groups in regions south of the Equator. These groups also had a narrower $\delta^{13}C$ range, as seen in less elongated ellipses. The two SAG populations formed their own cluster, with ellipses elongated along the $\delta^{13}C$ axis. The S-EQ groups had the largest calculated ellipse area (SEAc) and the largest variation along the $\delta^{13}C$ axis.

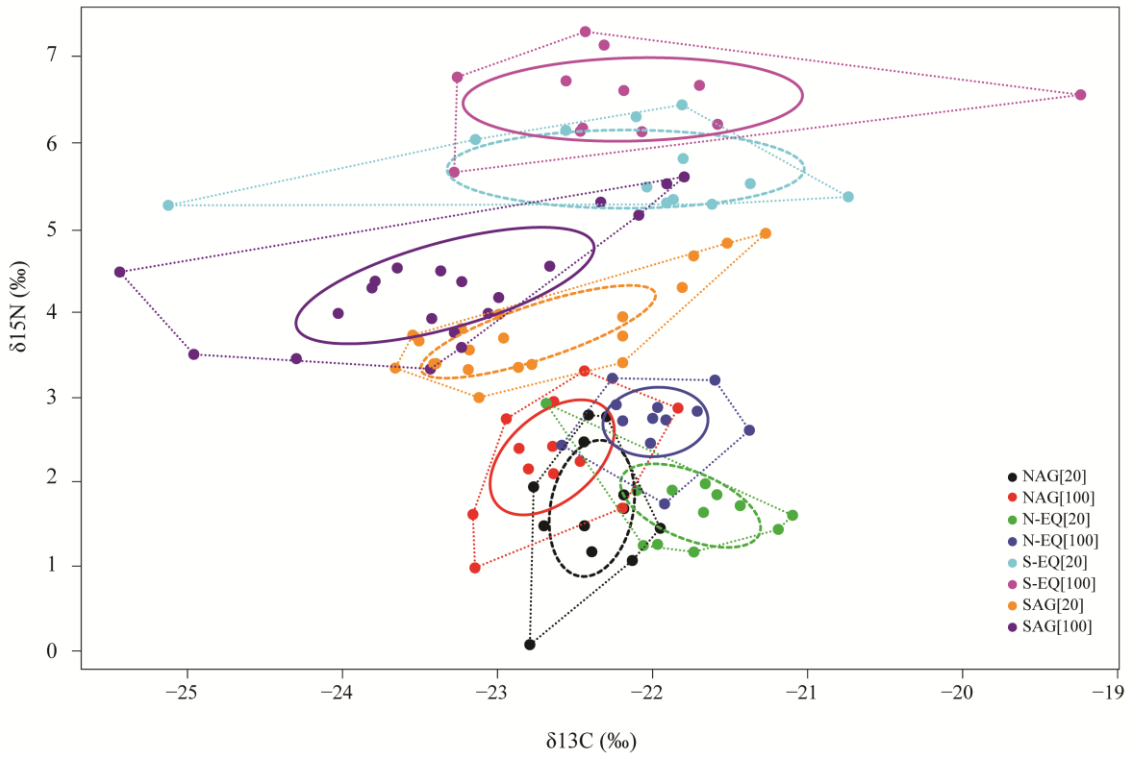


Figure 4.6 Isotopic niche space calculated using SIBER. Individual filter replicates were used and eight groups were identified before analysis (see text for more detail). These are shown in the legend. Coloured dashed lines show standard ellipse areas (SEAc) around $Micronet_{20}$, while coloured solid lines represent $Micronet_{100}$ samples (40% of data is included within an ellipse). Coloured dotted lines are convex hulls encompassing all datapoints.

4.4 Discussion

The stable isotope composition of microplankton populations from two size fractions was examined along a large spatial scale in the Atlantic Ocean in order to determine reference isotope data of mesozooplankton dietary sources for use in AMT24 trophic studies. The isotope values and biomass estimates of collected samples along the studied transect were significantly different among regions, reflecting the environmental conditions (e.g. oligotrophic vs. eutrophic waters) and biological processes (e.g. nitrogen fixation) occurring in the water column. Measurement of carbon and nitrogen biomass followed a pattern of lower biomass in the oligotrophic gyres when compared to the EQ region. The community composition in both size fractions in this study revealed a mixed community (Figure 4.2), containing both primary producers (e.g. *Trichodesmium*) and grazers (e.g. Dinoflagellates, copepod nauplii and Ciliates). The isotope measurements thus reflected an integrated signature across at least two trophic levels. The isotopic composition of *Pleuromamma* copepods was very similar to the Micronet stable isotope values and followed the same pattern across regions.

The variation in $\delta^{15}N$ of primary producers is a reflection of their nutrient sources and influenced by the biological processes involved in their uptake, e.g. N_2 fixation in subtropical gyres (Montoya, 2007). The low $\delta^{15}N$ values in samples from the NAG and EQ region, particularly at stations north of the Equator (1-3‰, Figure 4.5), indicate that N_2 fixation is important in food webs in this area. High abundances of *Trichodesmium* sp. were observed in stations sampled between 2° and 26° N (Figure 4.3B). This is one of the most important nitrogen fixers (diazotrophs) in the North Atlantic Ocean (Hood *et al.*, 2004; Davis *et al.*, 2006), with a latitudinal distribution between 0° and 20° N (Tyrrell *et al.*, 2003). The nitrogen isotope signature of diazotrophs has been measured to be between -1 to -2‰ (Montoya *et al.*, 2002). The observations of low $\delta^{15}N$ in the Micronet samples from this study are most likely explained by the high abundance of *Trichodesmium* in the samples. However, other diazotrophs (e.g. symbiotic cyanobacteria and Deltaproteobacteria) have also been shown to significantly contribute to nitrogen fixation in the open ocean (Zehr and Kudela, 2011; Grosskopf *et al.*, 2012) and could be contributing to the low $\delta^{15}N$ values observed here.

In stations south of the Equator along the AMT24 cruise track a sharp rise in $\delta^{15}N$ values was observed in the three stations in the EQ region, suggesting that the contribution from diazotrophs in the samples was smaller than in stations from the North Atlantic. The $\delta^{15}N$ values in samples from the SAG were significantly lower (between 3 and

5‰) than in the southern EQ samples (Figure 4.5), indicating the effect of nitrogen fixation in the food web. Although there are fewer studies that have investigated in detail the levels of nitrogen fixation in the South Atlantic Gyre than in the North Atlantic Gyre, there is evidence of both diazotroph presence (Sohm *et al.,* 2011; Fernandez *et al.*, 2014) and measurements of N_2-fixation rates, showing nitrogen fixation is an important process in this region (Moore *et al.*, 2009; Grosskopf *et al.*, 2012). It has been shown that lower levels of nitrogen fixation in the South Atlantic are due to reduced iron availability (Moore *et al.*, 2009). It is also likely that measurements of nitrogen fixation have been underestimated because diazotrophs other than *Trichodesmium* have been overlooked (Grosskopf *et al.*, 2012).

Trophic fractionation between the two size fractions ($\Delta^{15}N$) revealed a low increase in the $\delta^{15}N$ from $Micronet_{20}$ to $Micronet_{100}$ (average of 0.7‰), suggesting that microplankton from both size fractions grazed on similar food sources. This observation would, however, be better supported with the measurements of the 0.7-20 μm size fraction as a baseline. Although the isotopic composition of particulate material (between 0.7-20 μm) was not examined in this study, several previous studies have done this in the oligotrophic North Atlantic (Montoya *et al.*, 2002; Landrum *et al.*, 2011) as well as on the AMT scale (Fernandez *et al.*, 2014). These have shown that trends in the $\delta^{15}N$ values of particulate material reflect the isotopic composition of the nutrient nitrogen source (e.g. nitrate ranges between 3 and 6‰, while N_2 is 0‰). In areas in the ocean in which nitrogen fixation is the predominant process, low values (between -2 to 3‰) are therefore expected, while higher $\delta^{15}N$ values are expected where the main nutrient source of primary producers is nitrate (Montoya, 2007). In the North Atlantic measurements for similar size fractions are comparable to literature values (Montoya *et al.*, 2002; Fernandez *et al.*, 2014), showing evidence of nitrogen fixation influencing the food web. Although, measurements in this study in the South Atlantic are higher than ones from comparable studies (Fernandez *et al.*, 2014), they also reflect the observations made by others that nitrogen fixation has an impact on organisms futher in the trophic web.

A comparison of microplankton communities among different regions in the Atlantic Ocean revealed little change in community composition and estimated abundances according to the FlowCam technique. The main difference observed was in a higher contribution of primary producers (i.e. *Trichodesmium*) in samples from the North Atlantic, whereas the community composition in the southern samples consisted mostly of microzooplankton (i.e. dinoflagellates, ciliates, copepod nauplii). However, significant differences were observed in the carbon and nitrogen isotopic signatures, as well as

carbon and nitrogen biomass, among regions along the AMT24 cruise track. In the less productive oligotrophic gyre regions (NAG and SAG), lower biomass and lower $\delta^{13}C$ values were measured. The low biomass measurements had a similar pattern to *Synechococcus* abundances and oceanographic measurements described in Chapter 2. These were used for the determination of region boundaries along the AMT24 cruise track and the microplankton biomass measurements further complemented the region identification. Further interpretation of the $\delta^{13}C$ measurements made in this study was difficult because variation across trophic levels is very small and so fewer trophic studies have investigated it in detail. Differences in $\delta^{13}C$ have mainly been investigated across a larger spatial scale and between ocean basins (Rau *et al.,* 1982), between benthic and pelagic food webs (France, 1995) and looking at the effect of cell geometry on the carbon isotopic fractionation (Popp *et al.*, 1998).

Populations north of the Equator shared a similar isotope space and had smaller ellipse areas than those to the south, representing a narrower or more specialized isotopic niche. This observation is probably explained by a high contribution of new nitrogen from N_2 fixation. Populations from the southern EQ region and the SAG also clustered together in their isotope values. Both groups showed greater inter-population variability in $\delta^{13}C$ values than northern populations (Figure 4.6). This may reflect the differences in community composition of filtered samples, for example fewer primary producers (e.g. *Trichodesmium*) sampled in the South Atlantic stations, determining the isotopic composition of the microzooplankton or a greater variability in diet among different groups of microzooplankton. Both these factors could influence the isotopic niche space on the population level.

Comparing a set of *P. gracilis/piseki* and *P. borealis* isotope measurements to the Micronet baseline values showed that these copepods had a similar trophic position to that of the microplankton size fractions in this study. This was especially true in the EQ region and SAG, while animals measured in the NAG were slightly above the baseline. A wide range of $\delta^{15}N$ values was observed in this study, which suggests that the studied *Pleuromamma* copepods were feeding on a large size range of food. The results of this study suggest that it is likely that size fractions smaller than those sampled here (for example the 0.7-20 µm size fraction) form an large part of the *Pleuromamma* diet, particularly in the southern Atlantic where copepod $\delta^{15}N$ values were equal to the Micronet baseline. However, due to the small number of samples and stations measured more detailed conclusions of the trophic position are not possible at this stage. Other *Pleuromamma* species in the Atlantic have been shown to occupy various trophic positions

and characterised by a lack of storage lipids reflecting their active migrating lifestyle (Teuber *et al.*, 2014). The C/N ratios calculated for animals in this study were low, reflecting a high protein content probably linked to well-developed swimming musculature (Morris and Hopkins, 1983). They were, however, similar to those measured in other *Pleuromamma* species in the Atlantic (Teuber *et al.*, 2014) as well as to the Micronet C/N ratios.

In conclusion, this study provides a dataset of the isotopic and community composition of microplankton in the 20 – 180 μm size fraction, along the 2014 Atlantic Merdional Transect (AMT24). Observations showed latitudinal variation in isotopic compositions that could be explained by the variation in microplankton community composition (e.g. low $\delta^{15}N$ values in the North Atlantic due to a high abundance of *Trichodesmium* in the samples). The stable isotope measurements made for a set of *P.* gracilis/*piseki* and *P. borealis* samples suggested that these copepod species had a varied diet ranging from particulate material to microzooplankton. There was also an indication of differences among oceanic regions in the Atlantic but more samples need to be measured to make further conclusions. Our results provide useful reference data for future mesozooplankton studies, and baseline information for investigation of differences in microplankton community structure and isotopic composition across the Atlantic Basin.

Chapter 5

Diversity of bacteria associated with *Pleuromamma* copepods from the Atlantic ocean

5.1 Introduction

Copepods are the most abundant zooplankton group in the ocean (Bron *et al.*, 2011) and play a pivotal role in the transfer of energy through the pelagic trophic web (Banse, 1995). Additionally, they have a link to the microbial loop through their contribution to the pool of dissolved organic material (DOM) via various feeding mechanisms (Møller and Nielsen, 2001). Although many areas of biology have focused on studying copepods in various contexts, few studies have looked in detail at the relationship between copepods and their associated bacteria. There has recently been increased interest in characterising a direct association linking zooplankton and bacteria. Tang *et al.* (2010), for example, stressed the need of zooplankton bodies (both exterior and interior) and bacteria to be viewed as a dynamic system, not studied simply as two indirectly connected groups. Additionally, it has been hypothesised that copepods could represent a microhabitat or hotspot for microbial colonisation in an environment often characterised by its patchiness and limitation of nutrients (Tang, 2005).

Early studies of copepod-associated bacteria were based on culture-dependant methods and electron microscopy (e.g., Nagasawa and Nemoto, 1988; Hansen and Bech, 1996; Proctor, 1997) but these methods either missed much of the bacterial diversity or only allowed visualisation and not identification of the bacteria present. However, they did provide a first insight into the presence and function of bacteria associated with copepods.

A common taxon that has often been reported as associated with copepods is *Vibrio* spp., and the association is most likely external, with the *Vibrio* sp. attaching to chitinous exoskeletons of live copepods in the water-column (Huq *et al.*, 1983). Interestingly, when comparing the copepod-attached *Vibrio* to those living in the water column, most authors have found that the populations seemed to be independent of each other. This suggests that copepods represent a specific selective niche for chitin degrading *Vibrio* sp. (Heidelberg *et al.*, 2002; Shoemaker and Moisander, 2015). Other studies have found evidence of important biogeochemical processes, such as nitrogen-fixation, occurring through the interaction of copepods and diazotrophic microbes (Proctor, 1997; Braun *et*

al., 1999; Scavotto *et al*. 2015). In many cases, the internal bacterial community associated with copepods seems to be linked to or even introduced through the food source and its attached bacteria, with evidence showing that fecal pellets produced by copepods can be enriched in new bacteria compared to the original source (Møller *et al.*, 2007; Tang *et al.*, 2009; De Troch *et al.*, 2010).

With the development of molecular techniques, such as various sequencing methods, it has been possible to study the diversity of bacteria associated with copepods on a larger scale. A consistent finding of studies using molecular techniques (e.g. DGGE, cloning, Sanger and next-generation sequencing) to compare the copepod microbiome and the bacterial community of the surrounding water-column has been that they usually differ significantly (De Corte *et al*, 2014; Shoemaker and Moisander, 2015). This would indicate that copepods harbour a unique bacterial community. Most studies have also found that there is exchange between the copepod and water-column habitat (Bickel *et al.*, 2014; De Corte *et al*., 2014) and that bacterial community composition does not seem to differ among different copepod species (Gerdts *et al.*, 2013).

To date, the focus of copepod-bacteria interaction studies has been on freshwater species (Grossart *et al.*, 2009) and coastal or estuarine species (Hansen and Bech, 1996; Tang *et al.*, 2009). Several findings can be summarised from these diverse studies. Firstly, copepods seem to have a bacterial community associated with both their exterior (e.g. *Vibrio* colonisation) and interior (e.g. gut microbiota). Secondly, these associated bacterial communities are mostly different in species composition to that of the surrounding water-column. Thirdly, copepod guts can be seen as microbial hotspots for bacterial colonisation (Tang *et al*., 2010) but also representing a selective pressure for the community composition of copepod-associated bacteria (De Corte *et al.* 2014). Bacteria colonising copepod guts may be a key component in the biogeochemical cycling of organic matter, making nutrients accessible to the host and also to organisms in the surrounding water column (Freese and Schink, 2011).

In this study, the bacterial community composition associated with two copepod species, *Pleuromamma gracilis/piseki* and *P. borealis*, was investigated using an Illumina sequencing approach. Both species belong to an abundant and ubiquitous genus but have differing distributions with *P. gracilis/piseki* occurring in the oligotrophic and tropical Atlantic while *P. borealis* occurs predominately in the tropical Atlantic (Steur, 1932). The aim of the study was to characterise the community composition of *Pleuromamma*-associated bacteria across a large spatial scale in the Atlantic Ocean, more specifically

along a transect traversing major oceanic biomes, including two oligotrophic gyres as well as the Equatorial convergence region. Differences in bacterial community composition among the oceanic provinces and between the two copepod species were examined. Finding taxa that could potentially have a symbiotic relationship (e.g. gut microbiota) with the studied copepods was additionally of interest.

5.2 Methods

5.2.1 Sample collection and preparation

Adult individuals of *P. gracilis/piseki* (PLGP) and *P. borealis* (PLBO) were collected on the AMT24 research cruise that took place between September and November 2014 aboard the RRS James Clark Ross, as described in Chapter 2, Section 2.4.1.

Individual adult copepods from samples fixed in 2% paraformaldehyde and stored in 50% ethanol were chosen from selected stations along the transect (Table 5.1, Figure 5.1.). A total of 81 copepods were picked from bulk samples and the copepod species was identified based on external morphology where possible (details of host morphology can be found in Chapter 3, Figure 3.1). The copepod samples sequenced were a mix if *P. gracilis/piseki* and *P. borealis* and details on species ID can be found in Appendix B.1, Table B3. Sequencing of the mitochondrial COI gene for additional molecular identification of chosen individuals was done as described in Chapter 3 (Figure 3.2).

Table 5.1 Stations selected for next-generation sequencing in Chapter 5. Region refers to regions as defined in Chapter 2, Section 2.4 (North Atlantic Gyre – NAG, Equatorial Region – EQ, South Atlantic Gyre – SAG). N is number of individuals per station chosen for sequencing.

Cruise-Station	Latitude	Longitude	Date sampled	Region	N
ZPT_JR303_006A	31° 52.296 N	27° 33.151 W	03/10/2014	NAG	8
ZPT_JR303_007	27° 2.978 N	28° 52.317 W	05/10/2014	NAG	3
ZPT_JR303_008A	24° 37.307 N	29° 44.774 W	05/10/2014	NAG	5
ZPT_JR303_011	14° 12.596 N	27° 55.385 W	09/10/2014	EQ	9
ZPT_JR303_013	7° 17.2674 N	26° 29.459 W	11/10/2014	EQ	9
ZPT_JR303_015	0° 4.820 N	25° 52.620 W	13/10/2014	EQ	7
ZPT_JR303_016	3° 53.219 S	25° 1.280 W	14/10/2014	EQ	5
ZPT_JR303_017	7° 28.263 S	25° 2.212 W	15/10/2014	EQ	4
ZPT_JR303_018A	10° 26.230 S	25° 2.836 W	16/10/2014	SAG	9
ZPT_JR303_020	18° 18.986 S	25° 4.979 W	18/10/2014	SAG	8
ZPS_JR303_024A	30° 35.490 S	25° 48.288 W	22/10/2014	SAG	14

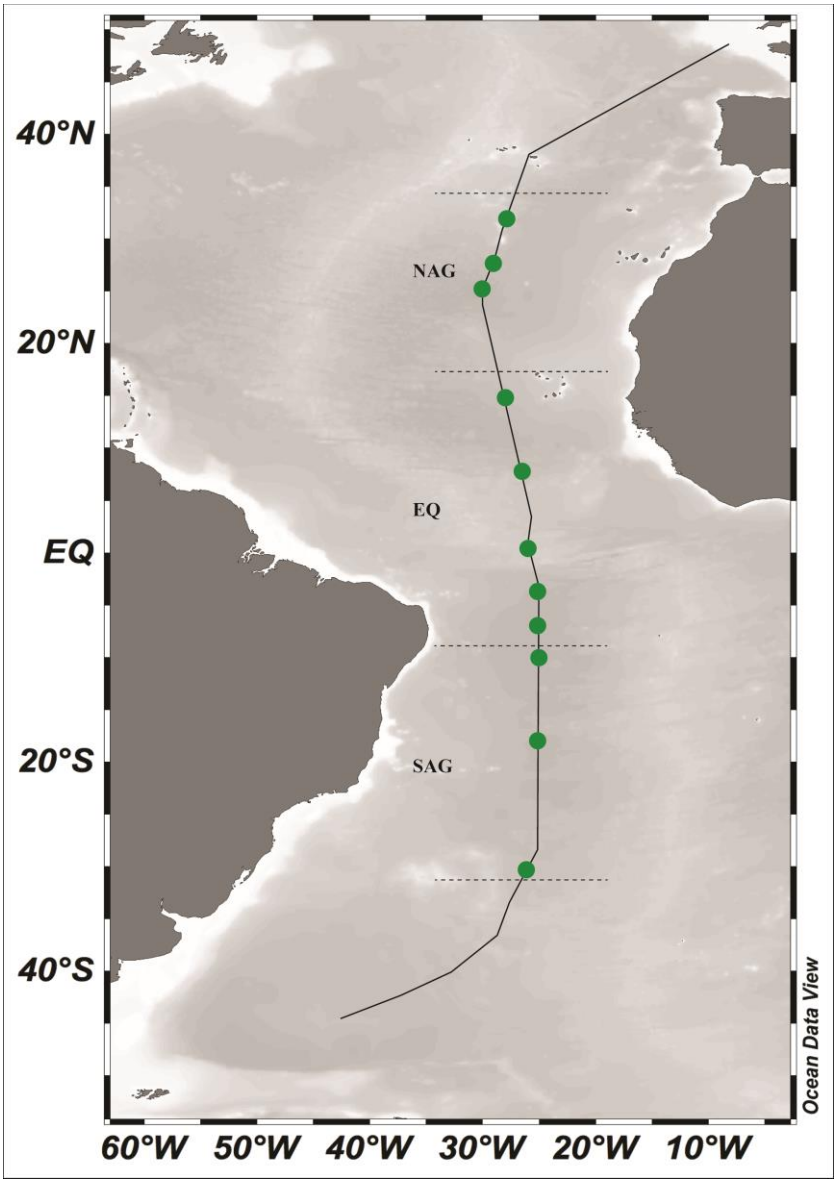


Figure 5.1 Map showing location of stations (green dots) along the AMT24 cruise track at which samples used in Chapter 5 were collected. Dashed lines show region boundaries (Northern Temperate – NT, North Atlantic Gyre – NAG, Equatorial Region – EQ, South Atlantic Gyre – SAG, Southern Temperate – ST).

5.2.2 Sequencing of bacterial 16S rRNA genes

In order to determine the bacterial community composition of individual PLGP and PLBO copepods, samples were sequenced using Illumina technology. Several sample preparation steps were performed at the Max-Planck Institute for Marine Microbiology (MPI-MM) prior to sending samples for library preparation and sequencing at the Max Planck-Genome-Centre (MPI-GC) in Cologne, Germany.

Firstly, DNA was extracted from copepod samples following a protocol modified from Zhou *et al*. (1996). All steps, unless indicated differently, were done at room temperature (RT) and all centrifugation was done at 13,000 rpm. Individual copepods were incubated at 37 °C for 6 days in a mix of 90 µl Extraction buffer (EB; 100 mM Tris/HCl (pH 8.0), 100 mM EDTA (pH 8.0), 100 mM Na/Phosphate (pH 8.0) and 1% CTAB) and 10 µl Proteinase K (ProK; 20 mg/ml; Macherey-Nagel, Germany). Additional ProK was added halfway through the incubation. This initial long incubation was necessary to ensure complete digestion of tissue in samples fixed with PFA. After incubation, 10 µl of 20% SDS was added and samples incubated for 2 hours at 65 °C on a thermo-shaker (Eppendorf, Germany), then centrifuged for 20 min and the liquid transferred to fresh tubes. This step was repeated with the addition of 90 µl EB and 10 µl 20% SDS, incubating for a further 20 min at 65 °C, centrifuging for 20 min then transferring the liquid phase to the liquid collected in the previous step. A Phenol-Chloroform-Isoamylalcohol (PCI; 25:24:1 v/v; Ambion, USA) extraction was performed by adding one volume of PCI to samples and incubating for 15 min, then centrifuging for 20 min. The upper aqueous phase was transferred into fresh tubes and a second extraction with Chloroform-Isoamylalcohol (CI; 24:1 v/v; Merck, Germany) was done. Again, one volume of CI was added to samples and centrifuged for 20 min, after which the upper aqueous phase was transferred to fresh tubes. The DNA was precipitated by adding 0.6 volume of 2-Propanol (Merck, Germany) and incubated for one hour. Samples were centrifuged for 30 min, after which the supernatant was carefully discarded, the pellet washed with 1 ml of 70% EtOH and centrifuged for a further 15 min. After carefully discarding the supernatant, the pellet was left to air-dry to insure all EtOH was removed. Finally, the DNA pellet was eluted overnight at 4 °C in 30 µl low-EDTA 1x TE buffer.

Secondly, the V3-V4 hypervariable region of the 16S rRNA gene was targeted for Illumina MiSeq sequencing using barcoded primers. Unique barcode identifier sequences (7 bp in length, with a minimum of 2bp difference) generated using *barcrawl* software (Frank, 2009) were added to both forward (319F; ACTCCTACGGGAGGCAGCAG) and reverse (806R; GGACTACHVGGGTWTCTAAT) primers. Additionally, heterogeneity spacers (ranging between 0 and 7 bp) were added to the primers following the barcode sequence. A total of 27 unique barcoded primer pairs were used in this study (see Appendix B, Tables B.1 and B.2 for sequence details). Primers that provide good overall phylum coverage and therefore reduce bias were chosen according to Klindworth *et al.* (2012).

Polymerase Chain Reaction (PCR) amplification of the 487 bp hypervariable region was done in triplicate for each of the 81 samples. Briefly, reactions were carried out in a total

volume of 20 µl using the Phusion Polymerase PCR Kit (NEB, USA) and following the manufacturers recommendations. The reaction mix contained: 1x HF PCR buffer, 0.2 mM of each dNTP, 3% DMSO, 2.5 µM of each primer, 0.02 U/µl Phusion DNA polymerase and between 1 – 1.5 µl DNA template. The temperature profile was a 30 s initial denaturation step at 98 °C followed by 30 cycles of 10 s denaturation at 98 °C, 30 s of annealing at 58 °C and 15 s elongation at 72 °C; the PCR was completed with a final 10 min extension step at 72 °C. A negative control PCR without template was done for each primer pair to ensure there was no unspecific amplification.

The PCR products were resolved on 1.4% agarose-TAE gels with a low mass DNA ladder (Invitrogen, Carlsbad, USA) and run for approximately 35 min at 100 V. Agarose gels were post stained for 20-30 min using Ethidium Bromide DNA stain and visualised under an UV light source. Amplified products of the correct length and a single band on the agarose gel were pooled by sample and directly purified using a QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). PCR products for which more than one band was amplified in a single sample were cut from the agarose gel at the correct length, pooled and purified using the QIAquick Gel Extraction Kit (QIAGEN). The concentration of the amplified PCR products was measured with a Quibit® 2.0 Fluorometer (Invitrogen, USA). Samples were then pooled in equimolar amounts to a final concentration of 100 ng per individual sample and a minimum of 1 µg DNA in the final pool. The 81 PCR amplicon samples were divided among three separate pools (grouped by the 27 barcoded primer combinations) and sent for TrueSeq DNA library preparation and sequencing of 300 bp paired end reads on the MiSeq 2500 (Illumina) sequencer at the MPI-GC in Cologne, Germany. A total of 1,000,000 reads were requested per library.

5.2.3 Bioinformatics analysis

The following steps were performed separately for each library. Returned sequences were checked for read quality with FastQC (Andrews, 2010). All data generated in the Illumina MiSeq runs were above Illumina's quality specifications (70% data >= Q30). Illumina adapters (PhiX) and bad quality sequences (qual < 2) were removed using the BBduk script. Forward and reverse sequences were then merged with no mismatches in overlapping sequences allowed using the BBmerge script. BBduk and BBMerge are part of the BBmap package freely available online (Bushnell B. – sourceforge.net). Finally, barcode trimming and splitting was done with mothur v.1.31.2 (Schloss *et al.*, 2009). Samples from

all three libraries were then merged and renamed to a file input compatible with Minimum Entropy Decomposition (MED) software (Eren *et al.*, 2014).

The MED software was used to resolve sequences into taxonomically unique and meaningful units. All sequences were trimmed to an equal length of 400 bp using an inbuilt MED script. The decomposition pipeline was then run using default settings, except for parameter M, which was set to 100 (recommended setting is N/10,000 or larger, where N is the number of reads). Minimum substantive abundance (M) is a parameter in the decomposition pipeline that determines the minimum number of unique sequences in a node for it to be kept in the analysis and subjected to Shannon entropy analysis (Eren *et al*., 2013).

Output from the MED pipeline was used for taxonomical classification of individual MED nodes by submitting a file containing node representative sequences to the online SINA aligner (Pruesse *et al*., 2012). Taxonomic paths to genus level were used for community composition identification. However the resolution power of MED is more powerful than currently available taxonomic classification, so unique nodes could obtain the same taxonomic classification at genus level.

5.2.4 Statistical analysis

All statistical analyses were done in R statistical software version 3.2.1 using the Community Ecology Package 'vegan' (Oksanen *et al.*, 2016). Additionally, statistical functions were modified from freely available R script deposited in the online GitHub repository (https://github.com/chassenr/NGS) were used for data analysis.

Alpha diversity of the community dataset (sequence count data) obtained with MED analysis was explored by calculating Hill numbers H0 (species richness), H1 (exponential Shannon entropy) and H2 (Inverse Simpson concentration) as implemented by Chao *et al.* (2014). Hill numbers are a family of diversity indices, mathematically unified through an exponent q and they incorporate both relative abundances and species richness of samples. They are expressed as the effective number of species (Chao *et al.*, 2014). Rarefaction curves were also calculated to check for adequate sequencing depth of samples (see Appendix B, Figure B.2 for details).

The raw community dataset output was filtered to remove rare nodes and singletons prior to beta diversity analysis. This was done by removing nodes with less than two sequences and present in less than 10% of the samples. Additionally, the count data was

converted to relative abundance data for further analysis. A Mantel test using a Bray-Curtis dissimilarity matrix (*vegdist* function in R package vegan) was used to test whether filtering the dataset affected the beta diversity pattern when compared to the original data.

Two distance-based methods were used to determine whether patterns could be seen in the community dataset that can be explained by assigning samples to explanatory variables (e.g. regions along the transect or copepod species). Hierarchical cluster analysis was done using the Bray-Curtis method (*hclust* and *vegdist* functions in R package vegan). Similar groups were determined based on the cluster diagram at 98% similarity. This was done without considering *a priori* knowledge of samples belonging to different regions or either copepod species. A graphic representation of differences in community composition among samples was generated with non-metric dimensional scaling (NMDS) plots (*metaMDS* function in R package vegan). Samples were coloured based on either region or copepod species.

One-way ANOSIM was used to test for significance among regions along the transect and between the two copepod species. Posthoc tests were performed on significant tests to determine significant comparisons using the Bray-Curtis method to generate a distance matrix and the 'fdr' (Benjamini and Hochberg, 1995) method for p-value correction.

5.2.5 Supplementary information – Appendix B

Appendix B contains tables and figures that provide information relevant to the copepod-associated bacterial community study but were not included in the main chapter.

Additionally, Greta Reintjes and colleagues (MPI-MM) conducted a study of the water-column bacterial community composition along a similar transect in 2012 (AMT22) using 454-pysosequencing. Water samples for sequencing were collected on 0.2 µm (free living) and 3 µm (attached) populations from 20 m depth at several stations along the transect (see AMT22 Cruise Report for more detail). DNA extracts were sequenced using 454-pyrosequencing technology and sequences analysed using the SilvaNGS pipeline. These data were used for comparison of the water-column and copepod bacterial communities.

5.3 Results

5.3.1 Sequence analysis results

Eighty-one samples were sequenced in three separate libraries resulting in a total of 4,558,168 reads. After strict merging, trimming and splitting steps, a total of 1,134,483 sequences across 79 samples were submitted to the MED pipeline. Two samples were removed from further analysis due to inadequate sequence quality. Quality filtering resulted in the removal of 136,350 sequences due to violation of criteria –M (minimum number of sequences in node) and –V (maximum variation allowed in each node). This resulted in 998,134 sequences partitioned among 4853 nodes (representing unique sequences). MED diagnostics can be found in Appendix B, Table B.3.

Filtering of the dataset to remove rare nodes with less than 2 reads in samples and present in less than 10% of samples resulted in a total of 566 nodes left. A Mantel test, comparing the original and filtered dataset, showed a significant level of correlation between the two datasets ($R = 0.967$, $p < 0.001$). This confirmed that the Beta diversity pattern was not affected by reducing the number of nodes and the reduced dataset was used for further Beta diversity analyses and taxonomic composition visualisation.

5.3.2 Diversity of Pleuromamma-associated bacteria along the AMT24 cruise track

Hill numbers, diversity indices mathematically unified through the exponent q, H0 (species richness), H1 (exponential Shannon entropy) and H2 (Inverse Simpson concentration) were calculated to examine the alpha diversity of the community dataset. They are plotted in Figure 5.2, showing values per station with an interquartile confidence interval. The relatively steep drop between the species richness (H0; numbers on average between 60-80) and Inverse Simpson Index (H2; numbers falling to approximately 20) indicated that a few sequences dominate the sequence diversity. Hill numbers H1 and H2 were very similar among most stations (ranging between 30-45 for H1 and 10-20 for H2) with overlapping confidence intervals. The species richness (H0) of these stations was more variable with values ranging from 25 to 85; however, the confidence intervals still overlapped. Two stations, ZPT-017 (EQ) and ZPT-007 (NAG), had Hill numbers above or below the rest, respectively. ZPT-007 had a species richness of 25 and the lowest H1 and H2, while ZPT-017 had the highest species richness at almost 120 and the highest H1 and

H2 compared to other stations. Diversity indices for individual samples are shown in Appendix B, Figure B.1. Rarefaction curves for most samples reached a plateau, indicating that a sufficient sequencing depth was reached (Figure B.2 in Appendix B).

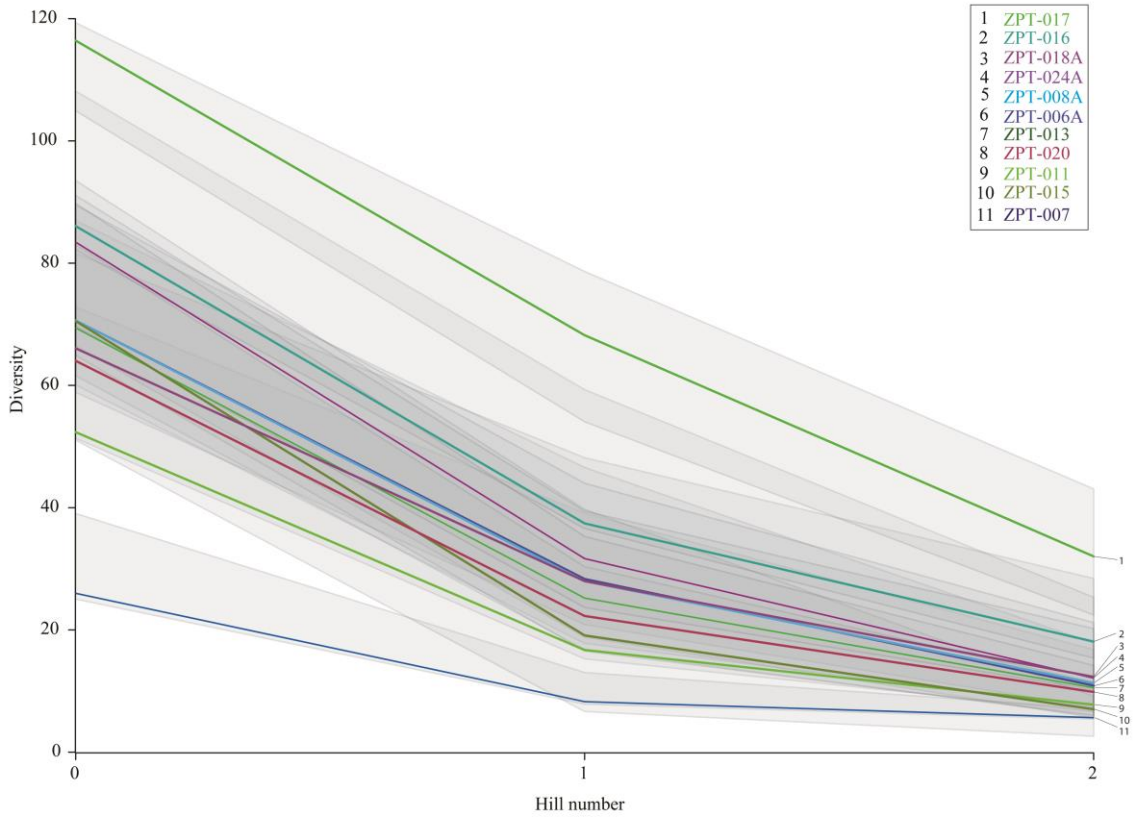


Figure 5.2 Alpha diversity calculated as Hill numbers (0-2). Each line, representing diversity per station, is shown with interquartile range confidence intervals plotted in grey. Stations plotted in blue colours were from the NAG region, green colours were from the EQ region and purple colours were from the SAG region.

Beta diversity analysis was done using Bray-Curtis dissimilarity matrix-based approaches. A cluster dendrogram of the pruned community dataset was plotted and five groups were identified at 98% dissimilarity level (Figure B.3, Appendix B). Samples were plotted using an NMDS plot with point colour corresponds to the region from which the sample came from (Figure 5.3). A similar plot with data points showing copepod species is in Appendix B (Figure B.4). Both analyses showed no clear grouping of samples by region or copepod species but rather a mixed distribution of all samples.

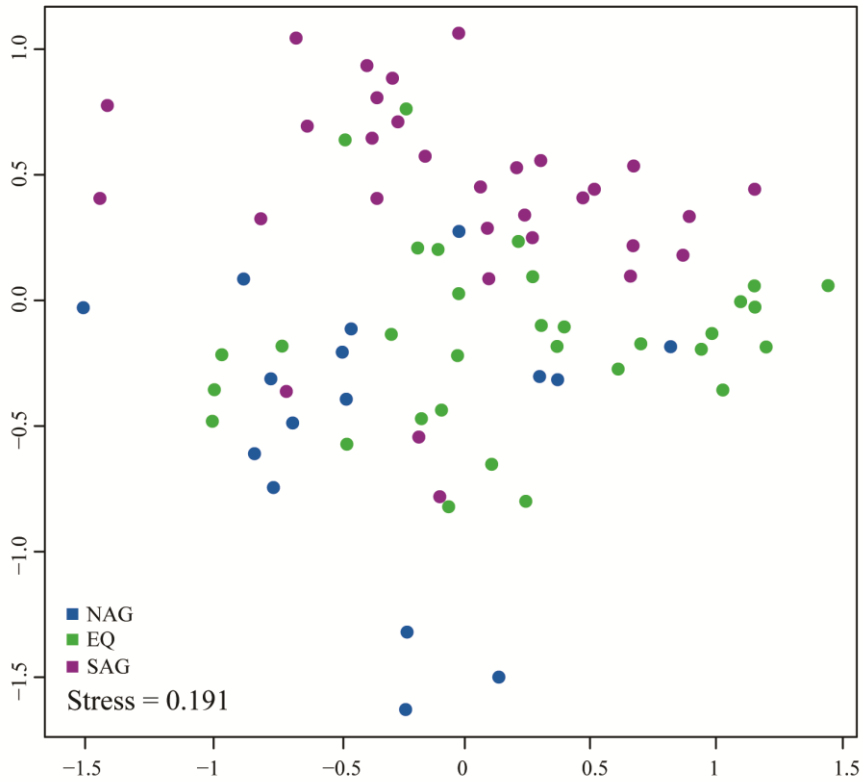


Figure 5.3 NMDS plot (using Bray-Curtis dissimilarity) of the filtered community dataset across 79 copepod samples. Colour of points corresponds to region along AMT24 cruise track. Blue: NAG, Green: EQ and Purple: SAG. The stress value of the ordination is given in the bottom left of the plot.

An ANOSIM approach was used to test for significant differences among regions along the AMT24 cruise track and between the two copepod species (PLGP and PLBO) studied here. No significant difference was found when the dataset was tested against copepod species as an explanatory variable ($R = 0.0006$, $p = 0.504$). However, when region was used as an explanatory variable, a significant difference was seen ($R = 0.207$, $p < 0.001$). Post hoc testing showed that all comparisons were significant (NAG-EQ: $R = 0.199$, $p < 0.01$; NAG-SAG: $R = 0.388$, $p < 0.01$; EQ-SAG: $R = 0.127$, $p < 0.01$).

5.3.3 Taxonomic composition of *Pleuromamma*-associated bacteria

Assigning taxonomic paths to MED nodes revealed a high diversity of taxa associated with *Pleuromamma*-copepods. The dataset, after filtering, contained sequence reads related to 12 different Phyla, 24 Classes and 125 Genera. The general pattern of taxa across samples and among regions was fairly consistent with no clear difference in community composition. Summaries of the most abundant phyla and classes are given in Figure 5.4 and Figure 5.5, respectively. Certain taxa were consistently present across all

samples. Between 20-80% of the relative abundance across samples was represented by the phylum Proteobacteria, among which the most abundant classes were Gammaproteobacteria (10-90%), Betaproteobacteria (0-80%), Alphaproteboacteria (5-20%) and Deltaproteobacteria (0-40%). Other dominant phyla included Cyanobacteria (2-85%), Firmicutes (5-20%, represented by the class Bacilli), Planctomycetes (5-40%, represented by Phycisphaerae), Actinobacteria (5-20%) and Bacteriodetes (5-20%). Other taxa, such as the classes Sphingobacteriia (Bacteriodetes; approximately 10%), Epsilonproteobacteria (Proteobacteria; 0-10%) and the phylum Verrucomicrobia (approximately 2%) were only present in low relative abundances in selected samples. Individual samples also had a high proportion of unclassified sequences.

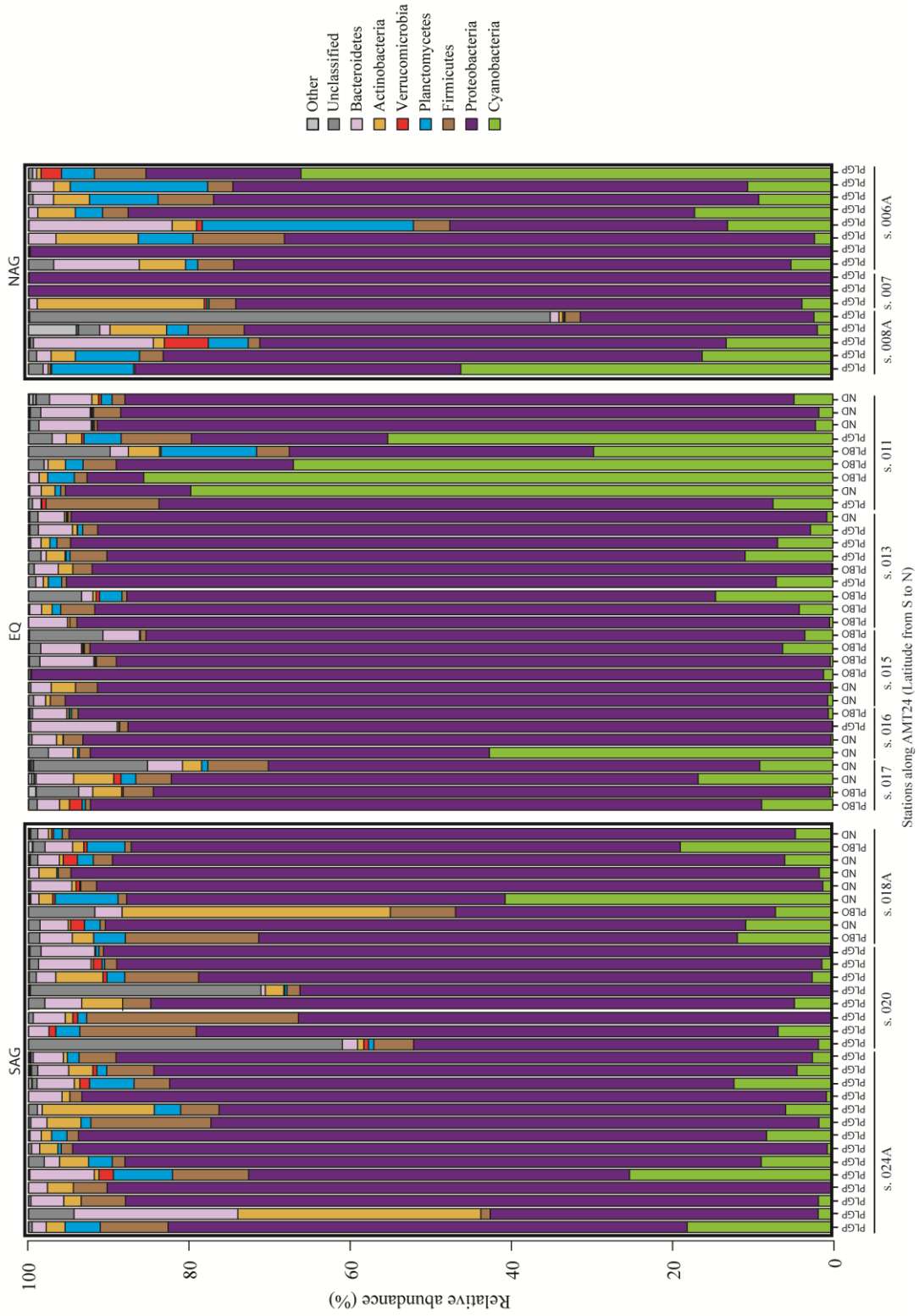


Figure 5.4 Barplot showing most abundant bacterial phyla per sample. Each bar represents an individual copepod sample, with the species ID indicated below (PLGP – *P. gracilis/piseki*, PLBO – *P. borealis*, ND – ID not possible). Bars are grouped by region, with station labels shown below samples. Stations are ordered from South to North.

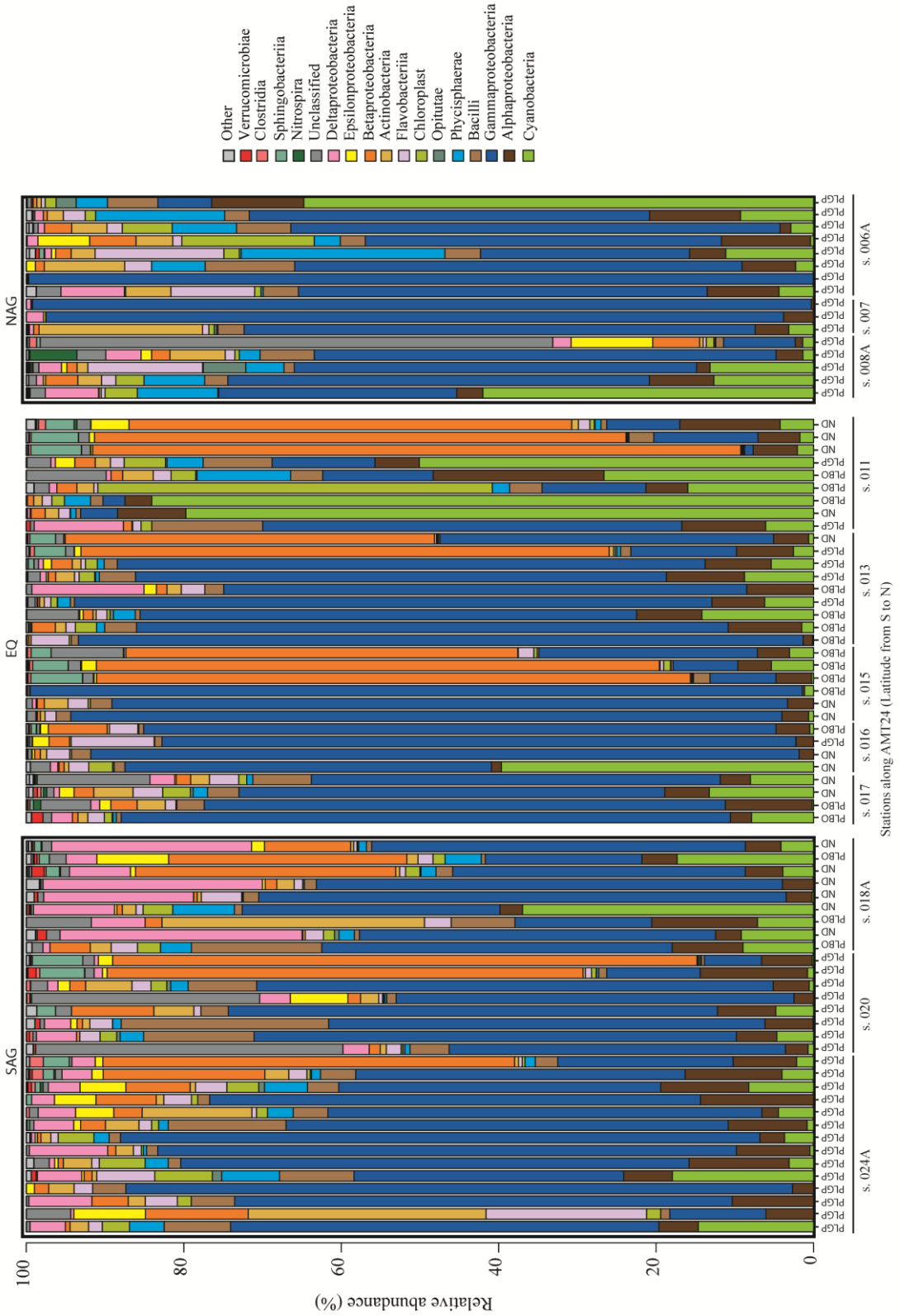


Figure 5.5 Barplot showing most abundant bacterial classes per sample. Each bar represents an individual copepod sample, with the species ID indicated below (PLGP – *P. gracilis/piseki*, PLBO – *P. borealis*, ND – ID not possible). Bars are grouped by region, with station labels shown below samples. Stations are ordered from South to North.

The high-resolution power of the MED pipeline allows sequence reads to be distinguished on the nucleotide level, thus revealing higher sequence diversity within the genus level. The relative abundances of the first 50 nodes containing the highest number of sequences are plotted in Figure 5.6. Only these nodes are shown because the relative abundance of other nodes across samples was on average below 0.5% and therefore considered less informative. The occurrence of individual nodes across samples plotted against their average relative abundance across samples is shown in Figure 5.7. The plot was divided into four quadrants by splitting the axes in half (i.e. at sample 39 and 5% average relative abundance) thus representing the contribution of individual nodes to the dataset. Three nodes were present in more than 50% of the samples and had an average relative abundance above 5%. They fell into Quadrant II (QII). Quadrants III (QIII) and IV (QIV) contained nodes that had a low average relative abundance but either low (9.5% of nodes) or high (90% of nodes) occurrence rate, respectively.

The three nodes with the highest relative abundance across samples present in QII (Figures 5.6 and 5.7) were most closely related to the betaproteobacterial genus *Limnobacter* (node 12892) and the gamaproteobacterial genera *Marinobacter* (node 22788) and *Alteromonas* (node 28683).

The highest number of abundant and frequent nodes belonged to Gammaproteobacteria (Figure 5.6). The genus *Vibrio* had a high diversity, with eight node representatives that were consistently present across samples. Other gammaproteobacterial nodes that had a high occurrence rate and were evenly distributed across samples were related to *Alteromonas* (2 nodes), *Pseudoalteromonas* (3 nodes), *Photobacterium* (5 nodes), *Marinobacter* (3 nodes) and *Moraxella*, *Alcanivorax* and *Pseudomonas* with one node each.

Several other nodes were also consistently present across samples, albeit in lower relative abundances (Figure 5.6). Such examples include: node 12763 (related to the genus *Sediminibacterium*, Sphingobacteriia, Bacteriodetes), nodes 24522, 13995 and 24506 (alpharoteobacterial genera *Methylobacterium, Thalassospira* and *Sphingomonas,* respectively), as well as two nodes related to *Micrococcus*, Actinobacteria (nodes 22247 and 22241) and three Bacilli, Firmicutes related nodes (node 29042 – *Staphylococcus,* node 27475 – *Sterptococcus* and node 23623 – *Paracoccus*).

A number of nodes, on the other hand, had a differential occurrence among stations or regions, with some nodes exhibiting higher relative abundances in only a few samples

(Figure 5.6). Two nodes related to deltaproteobacterial genera *Peredibacter* (node 21682) and *Halobacteriovorax* (node 22916), for example, were only present in higher relative abundances in samples from the gyre regions. Node 21682 had a relative abundance between 10-30% at station ZPT-018 (SAG) and around 5% at stations ZPT-008A and ZPT-007 (NAG). Another such node, with higher relative abundances in the gyres, as well as the only representative of the Epsilonproteobacteria among the most abundant nodes, was node 28935 related to the genus *Sulfurovum*. Similarly, two Phycispaerae (Planctomycetes) nodes (20493 and 25627) were more present in both gyre regions. Nodes with higher relative abundance and occurrence in the EQ region included several *Synechococcus*-related nodes and node 22879, related to *Mesoflavibacter* (Flavobacteriia, Bacteriodetes).

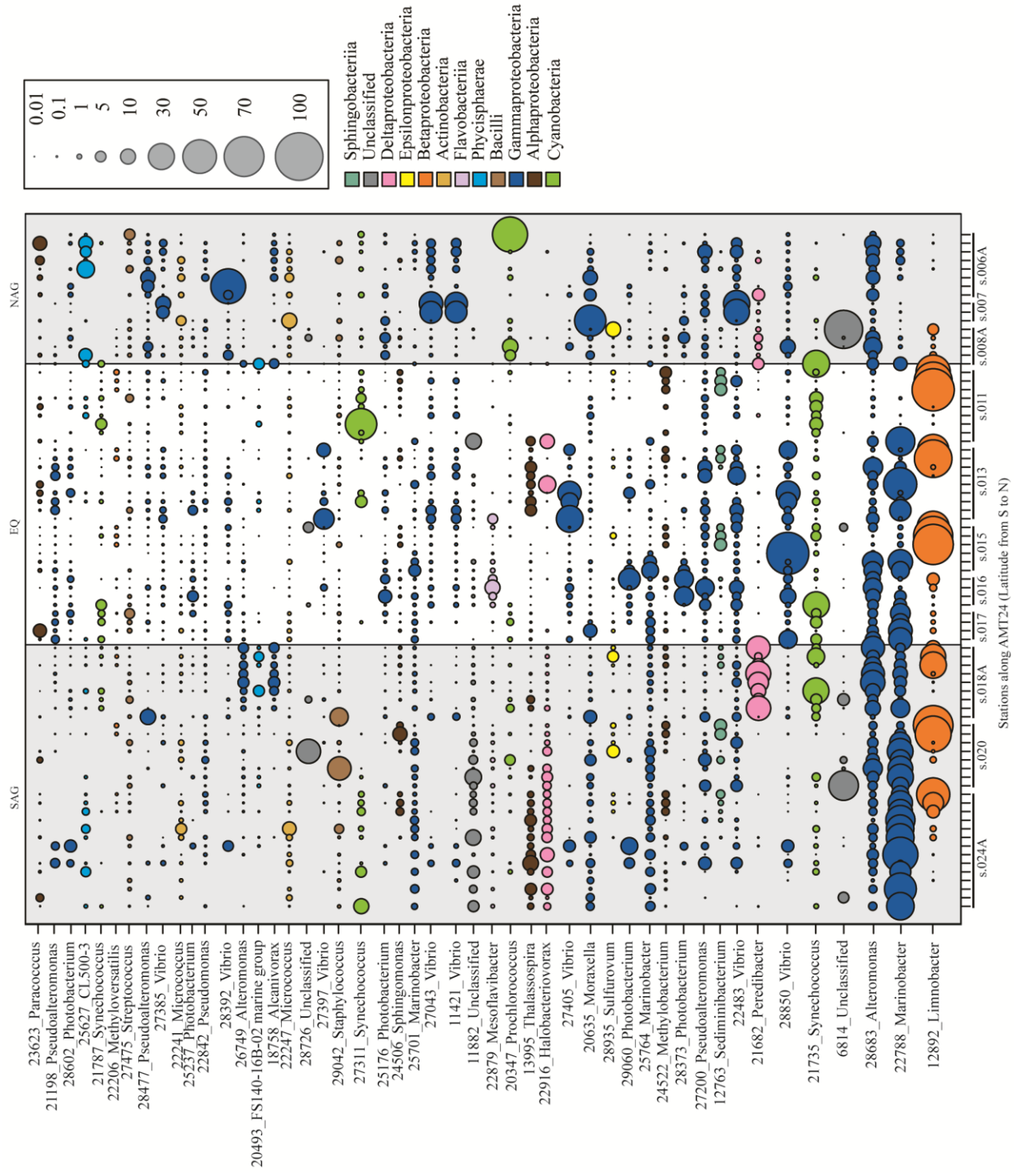


Figure 5.6 Relative abundances of 50 nodes containing the highest number of sequences. Each row represents an individual node and column show sequences obtained from individual copepods. Stations are labelled at the bottom, from South to North. Colours correspond to the classes shown in the legend in the bottom right of the figure, while the size of the circles represents the relative abundances in individual samples.

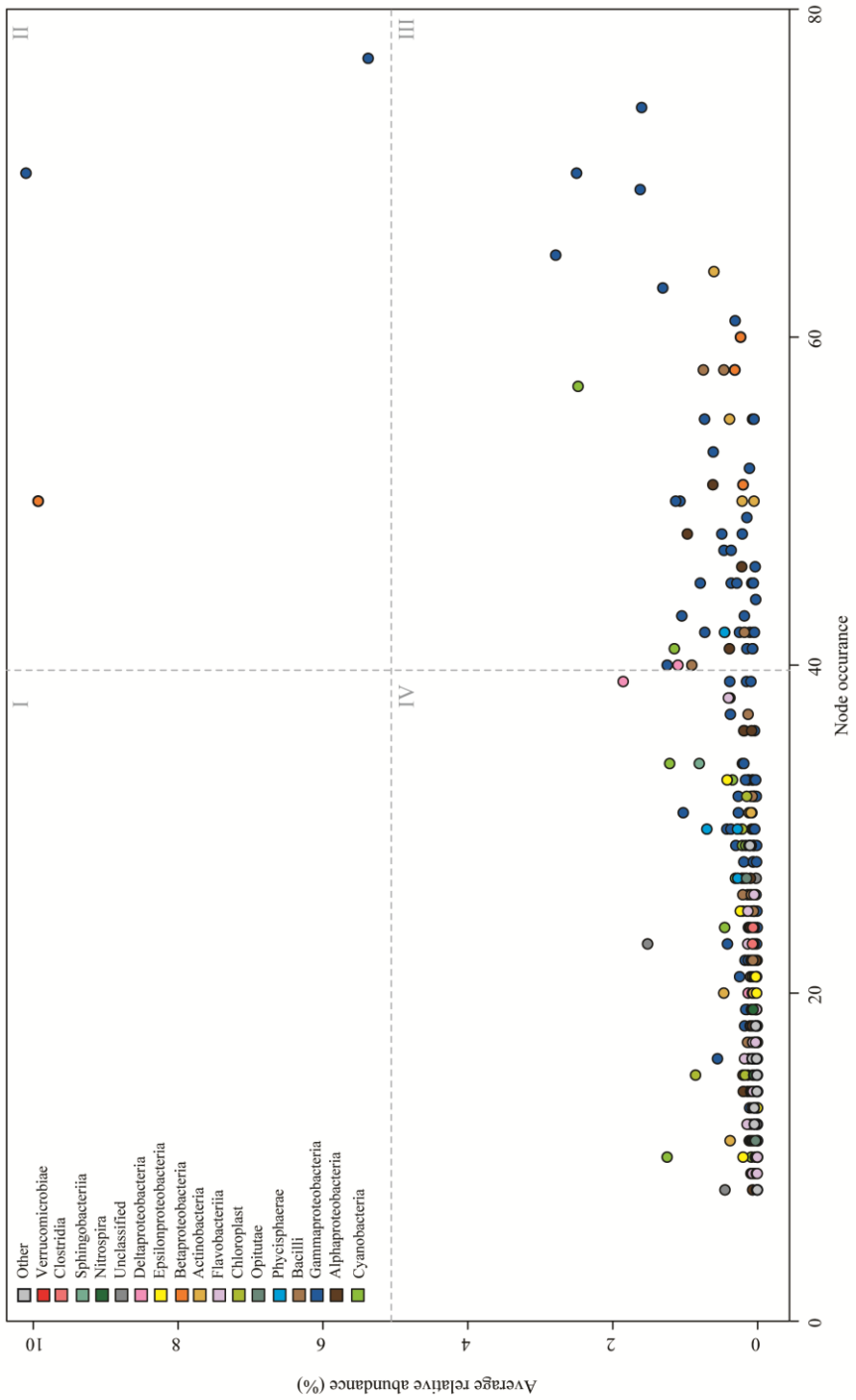


Figure 5.7 Plot showing average relative abundance and number of samples in which an individual node occurs. The colour of dots corresponds to the classes shown in the legend in the top left corner. The plot has been divided into four quadrants that can describe the contribution of nodes to the overall community composition. QI: nodes are present in high average relative abundance and low occurrence. QII: nodes are present in both high average relative abundance and high occurrence. QIII: nodes occur in more than 50% of samples but at low average relative abundances. QIV: nodes have both a low occurrence and low average relative abundance.

5.3.4 Comparison between *Pleuromamma*-associated and water-column bacterial communities

The bacterial community composition was compared between two water-column fractions collected during the AMT22 research cruise (collection and sequencing done by G. Reintjes and colleagues from MPI-MM) and the copepod-associated bacteria from this study. An NMDS plot with each dot representing an individual sample and point colour corresponds to the bacterial size fraction (i.e. free-living, attached and copepod-associated) is shown in Figure 5.8B. The three groups clustered separately from each other with more variability seen in the coppod-associated samples. Testing for significance using an ANOSIM approach showed a significant difference ($R = 0.158$, $p < 0.02$) and post hoc testing showed that only the comparisons between free-living vs. attached samples and free-living vs. copepod-associated samples were significant (free-attached: $R = 0.999$, $p < 0.01$; free-copepod: $R = 0.345$, $p < 0.01$). The comparison of attached vs. copepod-associated samples was not significant (attached-copepod: $R = -0.025$, $p = 0.6$). Figure 5.9 shows the comparison between the three bacterial community profiles at the bacterial class level. The general profiles differed among size fractions, with only some taxa (e.g. Gamma-, Alpha-proteobacteria and Cyanobacteria) found in relatively high abundance across communities. However, when these communities were compared at genus level, distinctly different profiles were observed (Figure 5.10). The most commonly occurring taxa in the copepod associated community, e.g. betaproteobacterial *Limnobacter*, gammaproteobacterial *Vibrio*, *Marinobacter* and *Alteromonas*, and epsilonproteobacterial *Sulfuromonas* were not among the most common taxa found in either water-column community.

Additionally, figures 5.9C and 5.10C could reveal more pronounced differences among the regions that could not be observed when looking at individual samples. For example, there was a difference in the relative abundance of Betaproteobacteria-related sequences in the NAG compared to the EQ and SAG. The NAG also seemed to have a higher proportion of unknown sequences.

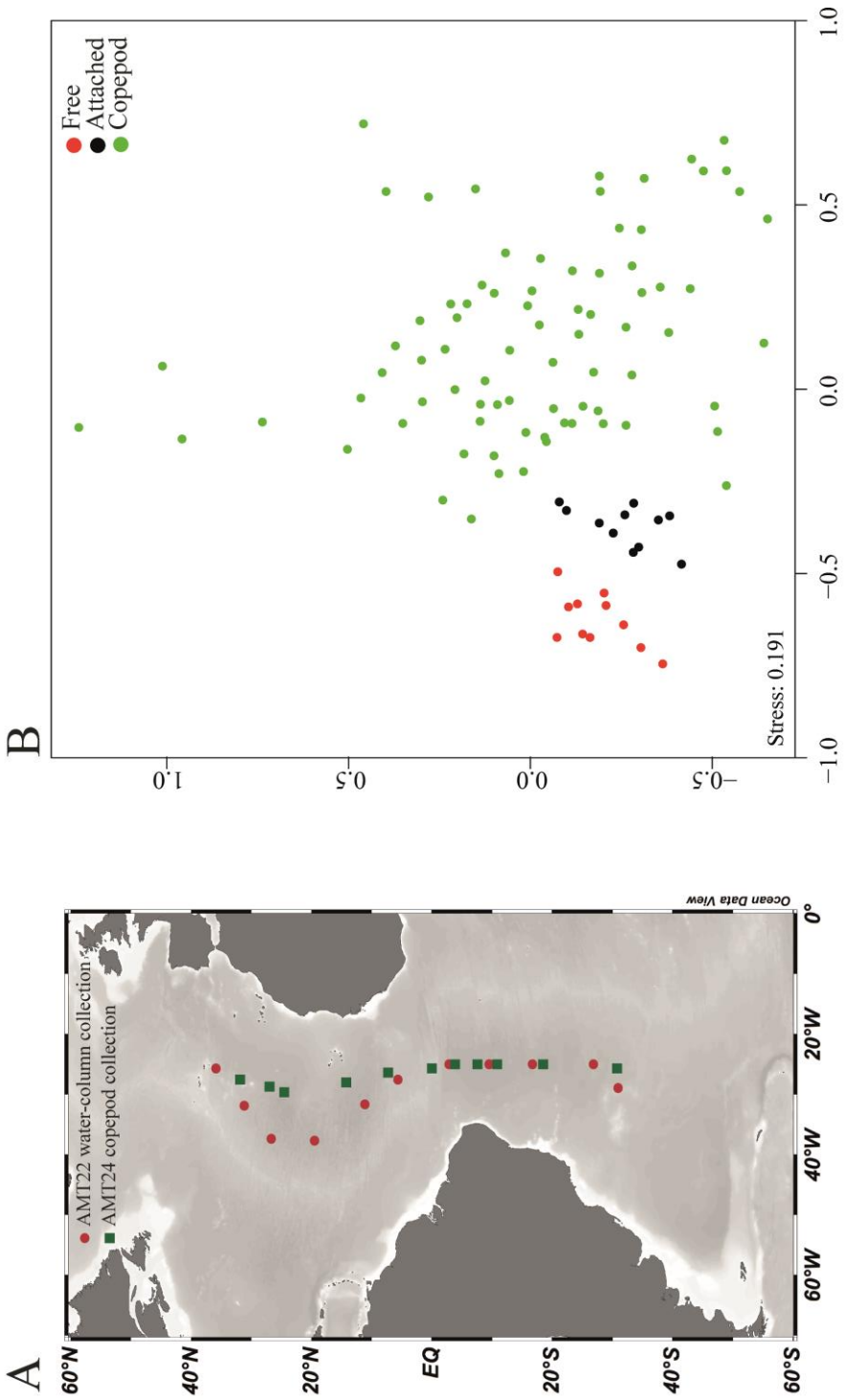


Figure 5.8 NMDS plot (using Bray-Curtis dissimilarity) of dataset with 79 copepod samples and both water-column size fractions at class level. Colour of points corresponds to dataset. Red: Free-living water-column size fraction (0.2-3 µm), Black: Attached water-column size fraction (3-10 µm)and Green: copepod samples. The stress value of the ordination is given in the bottom left of the plot.

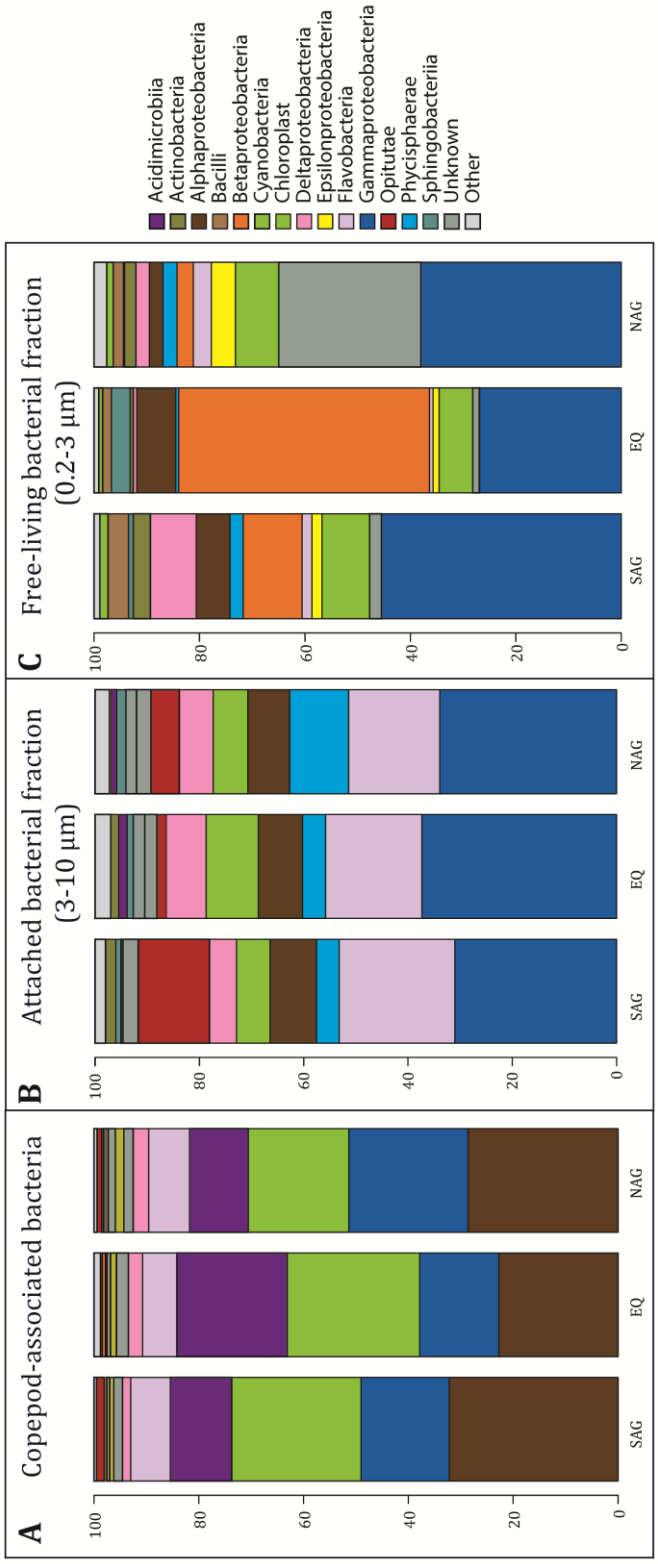


Figure 5.9 Comparison of bacterial community profiles among free-living **(A)** and particle attached **(B)** water-column bacteria collected on AMT22 and copepod-associated bacteria **(C)** from samples collected on AMT24. The barplots show cumulative profiles of bacterial classes per region (SAG – South Atlantic Gyre, EQ – Equatorial Region and NAG – North Atlantic Gyre).

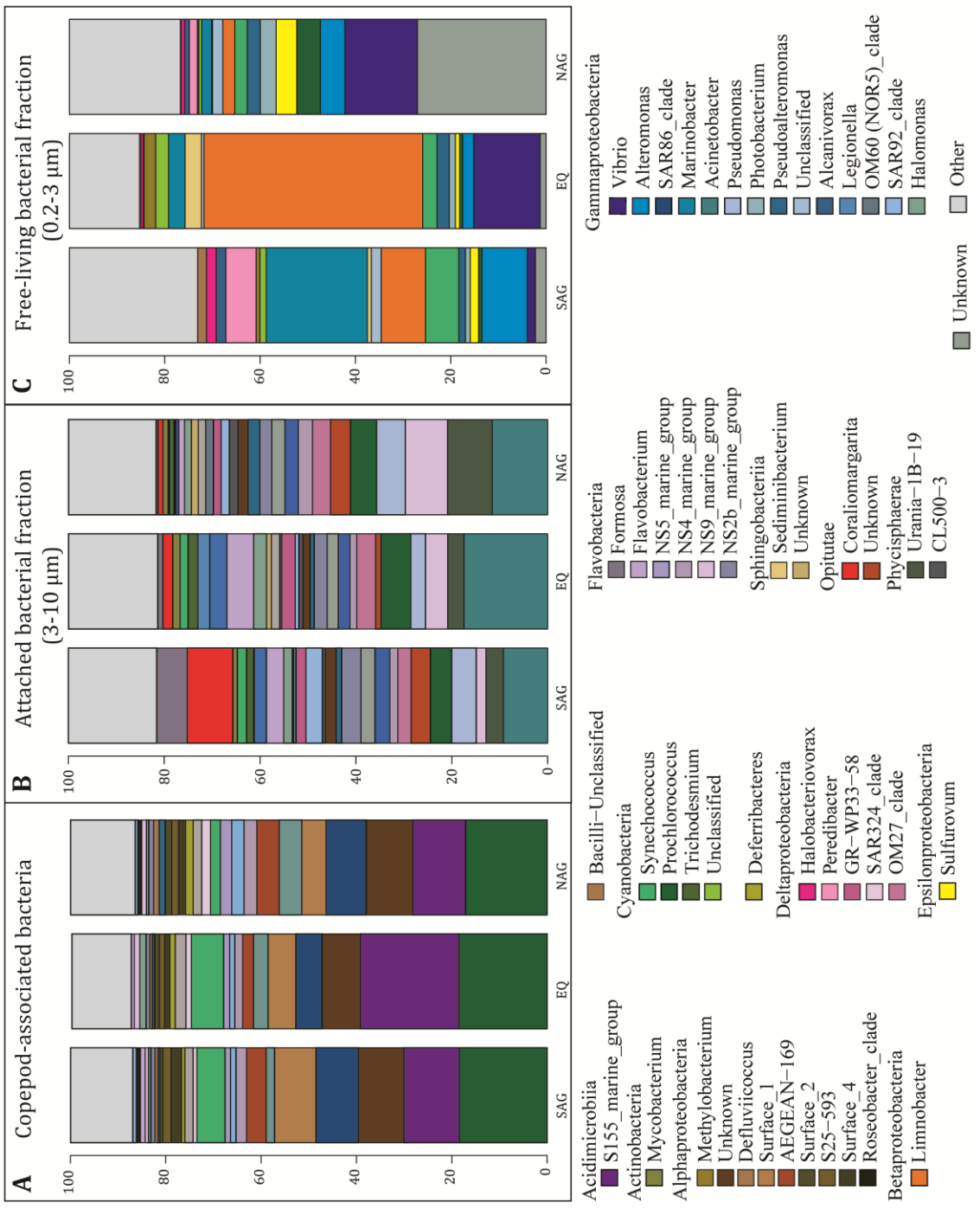


Figure 5.10 Comparison of bacterial community profiles among free-living (**A**) and particle attached (**B**) water-column bacteria collected on AMT22 and copepod-associated bacteria (**C**) from samples collected on AMT24. The barplots show cumulative profiles of bacterial genera per region (SAG – South Atlantic Gyre, EQ – Equatorial Region and NAG – North Atlantic Gyre).

5.4 Discussion

In this study the bacteria associated with *P. gracilis/piseki* and *P. borealis* copepods were investigated along the AMT24 cruise track. The high resolution this study offers on both a geographical spatial scale and in the number of individuals sampled is a significant contribution to the study of copepod-associated bacteria. Using Illumina sequencing of the 16S rRNA gene amplified from individual copepod DNA extracts, a high diversity of taxa associated with both copepod species was found. Nevertheless, only a small number of MED nodes (unique sequences) dominated the community composition across samples.

One of the most interesting nodes found in this study was related to the betaproteobacterial genus *Limnobacter* (node 12898). It had a high relative abundance in a number of samples from all regions (Figure 5.6), as well as the highest contribution across samples (Figure 5.7). Additionally, it was the only node related to Betaproteobacteria found in such a high relative abundance. This can also been seen when comparing Figures 5.9C and 5.10C, where the relative abundances of Betaproteobacteria and *Limnobacter* are very similar. The genus has been described as rod shaped, heterotrophic and strictly aerobic, as well as having the ability to oxidise thiosulfate to sulfate (Spring *et al.*, 2001; Lu *et al.*, 2011). Although the genus was originally isolated from freshwater sediments and volcanic deposits (Spring *et al.,* 2001), it has also been found in association with other marine organisms, such as the *Nematostella* sea anemone (Har *et al.*, 2015). The other most abundant nodes were related to gammaproteobacterial genera *Marinobacter, Alteromonas* and *Pseudolateromonas*. These genera, along with *Limnobacter,* have been reported to live in association with various diatom and dinoflagellate species (Amin *et al.,* 2009; 2012; 2015). All the gammaproteobacterial genera have been found directly associated either with copepod guts (Hansen and Bech, 1996; de Troch *et al.*, 2010; Tang *et al.*, 2011) or in association with other crustaceans (Givens *et al.*, 2013; Zwirglmaier *et al.*, 2015). While the Gammaproteobacteria were consistently found in the water-column dataset, *Limnobacter* was absent (Figure 5.9 and Figure 5.10).

Among the gammaproteobacterial nodes with high relative abundances were those related to the *Vibrio* genus. As discussed in the introduction of this chapter, they have been previously found associated with copepods and are hypothesised to most likely have an external interaction, attaching to the exoskeleton of copepods (Heidelberg *et al.*, 2002). The interactions, however, can range from beneficial through commensal to parasitic and many coastal copepod species can be carriers of diseases (Huq *et al*., 1983). Several nodes

found in the copepod-associated bacterial community were related to *Photobacterium* (*Vibrionaceae)*. Both of these gammaproteobacterial genera contain species that have been described to have luminescence properties (Haddock *et al.*, 2010). Bioluminescence has previously been observed in *Pleuromamma* copepods (Bannister and Herring, 1989), therefore it could be hypothesised that these taxa might have a luminescence role in the interaction.

Other nodes also contributed to the overall bacterial community, albeit in lower relative abundances. For example, the node related to the epsilonproteobacterial genus *Sulfurovum*, which is well characterised as a beneficial symbiont in many animal systems, providing its host nutrition through sulfur-oxidation and inorganic carbon fixation (Sanders *et al.*, 2013). Two nodes, belonging to distantly related genera (alphaproteobacterial *Methylobacterium* and betaproteobacterial *Methyloversatilis*) that can grow on single-carbon compounds (Lindstrom, 2001) were also associated with the copepod-bacterial community in this study. These are all nodes related to taxa with known and well-described activity. Others, however, could be equally important (e.g. nodes belonging to Sphingobacteriia or Opitutae), but speculation about their function is more difficult. An interesting lifestyle, however, has been described for two deltaproteobacterial nodes that appear more abundantly in the SAG region. They were affiliated with the Bacteriovoracaceae, more specifically the genera *Halobacteriovorax* and *Peredibacter*. They belong to a family of predatory bacteria (Davidov and Jurkevitch, 2004), recently found to have an important impact on controlling populations of other bacteria in estuarine environments (Williams *et al.,* 2015).

The diversity of the associated bacterial community did not differ significantly between the copepod species. Furthermore, very little difference was observed when comparing the whole community composition across oceanic biomes. The overall species richness (number of nodes in samples) of the samples fluctuated among stations (Figure 5.2 and Figure B.1 in Appendix B), which could affect the grouping of samples. Although the ANOSIM test with region as an explanatory variable gave a significant p-value, the R-value was very low ($R < 0.25$), which suggested that the groups were barely separable (Ramette, 2007). Oceanic biomes in the Atlantic have distinct physical and chemical characteristics (Maranon *et al.*, 2000 and Chapter 2 of this thesis), which would be expected to have an effect on the food source and water-column bacteria. Both of these are factors contributing to the community composition of copepod-associated bacteria, the former affecting the gut bacterial community and the latter providing a seed population for copepod colonisation (Tang *et al*., 2010). However, data obtained in this study about the part of the potential

copepod food community composition (i.e. microplankton without the size fraction below 20 µm, Chapter 4) and the AMT22 sequencing of water-column bacteria (data courtesy of G. Reintjes) found very little difference in community composition along each transect for both microplankton and water-column bacteria. These findings support observations regarding the uniformity of the *Pleuromamma*-associated bacterial communities along the transect. The similarity between the *P. gracilis/piseki* and *P. borealis* bacterial communities could probably be explained by the similar external and internal (i.e. gut) morphology of the host and observations of similar behavioural and trophic ecology (Steur, 1932). Indeed, several distantly related *Drosophila* sp. populations have been found to harbour very similar gut bacterial communities, especially among populations with the same diet and habitat preferences (Chandler *et al.,* 2011).

When looking at the community composition on a taxonomic level (Figures 5.4 and 5.5), the contribution of the dominant taxa was mostly uniform across samples (e.g. Gammaproteobacteria, Alphaproteobacteria and Cyanobacteria), with some exceptions that were more variable across samples. The presence of autotrophic Cyanobacteria and Chloroplast related sequences could indicate ingestion of primary producers since these groups are unlikely to form associations with the copepods and it is possible they had a full gut at the time of sampling. It has been hypothesised that copepods, including *Pleuromamma piseki/gracilis*, can actively ingest autotrophic picoplankton as part of aggregates or when they are in association with microplankton (Wilson and Steinberg, 2010).

The most pronounced difference when comparing different regions was seen in the Betaproteobacteria related nodes (Figure 5.8C). The relative abundance was clearly higher in the EQ when compared to both gyres and higher in the SAG when compared to the NAG. Other less dominant taxa only dominated certain samples or regions, such as Deltaproteobacteria in the SAG. Variation on an individual copepod level might have affected the overall comparison and these variations could be attributed to differences in the individual copepod's physiological state, fitness and whether they had a full or empty gut (De Corte *et al.*, 2014). Another explanation could be found in the seasonality between the regions, since samples were collected in the boreal autumn (NAG) and austral spring (SAG). Seasonality has been shown to directly affect both the standing stock and distribution of heterotrophic bacteria as well as phytoplankton in the Atlantic Ocean (Zubkov *et al.*, 2000). It is thus possible that changes in the picoplankton community could be influencing the copepod-associated bacterial community.

To date there have been only two other studies (De Corte *et al.*, 2014; Shoemaker and Moisander, 2015) using high-throughput sequencing techniques to study the copepod-associated bacterial diversity. Both studies focused on diverse groups of several copepod species and orders in the North Atlantic Ocean. Similar to this study, they found that the copepod-associated bacterial communities differed from the water-column bacterial communities and hypothesised the copepod interior and exterior as a specific niche. However, neither study was able to show highly specific and recurrent associations due to small individual copepod sample sizes (Shoemaker and Moisander, 2015) or pooling of individual at family level prior to sequencing (De Corte *et al.*, 2014). Some similarities could be found between the hypothesised community associated with their studies and this one. For example, Alpha-, Gamma- and Delta-proteobacteria were found in all studies, with similarities also at the genus level. Different betaproteobacterial genera were detected and these Betaproteobacteria were not found in the surrounding environment. However, they are not closely related to the Betaproteobacteria from this study, but rather belong to different families. Additionally, they were not found in such consistently high relative abundances.

It is important to note that the community composition of copepod-associated bacteria obtained in this study, as with the two other studies in which next-generation sequencing technologies were used (De Corte *et al.*, 2014; Shoemaker and Moisander, 2015), are based on short fragments (approximately 400 bp) of the 16S rRNA gene. Although this sequence length is sufficient for determination of sequence taxonomy and diversity, it does not allow more in-depth phylogenetic analysis of closely related taxa, such as full-length 16S rRNA sequences would. So although data obtained here did not show differences at higher taxonomic levels between copepod species and regions, it is possible that differences exist at the bacterial species or strain level. This was the case in ancient, well-established nematode and sulfur-oxidising bacteria symbiosis (*Candidatus* Thiosymbion), in which the phylogenies of several host species showed tight coevolution with their symbiont strains (Zimmermann and Wentrup *et al.*, 2016) but only at a full-length gene level.

In conclusion, the results of this study show a core copepod-associated bacterial community that is diverse at the phylum level, yet within those phyla represented by only a few genera. There is also likely a high metabolic diversity within this community. Although many taxa associated with the copepods were also found in the water-column dataset, we found a highly abundant betaproteobacterial genus *Limnobacter* that was unique to the copepod-associated bacterial community. The striking absence of

Betaproteobacteria in the water-column, the persistent high relative abundance of an identical sequence across the Atlantic Ocean suggests that they are part of a resident population. The results should however be viewed in a descriptive manner and more in depth analysis is needed for inference on the activity and type of association between the host and bacteria, for example using metagenomic and metranscriptomic approaches. Additionally, it is not possible to determine whether the bacteria form an internal or external association. However, this study does provide essential information for further work on *P. gracilis/piseki* and *P. borealis* bacterial microbiota and gives a first insight into the ecology and diversity of the copepod holobiont.

Chapter 6

Specific Betaproteobacteria associated with the gut of *Pleuromamma* copepods

6.1 Introduction

It is becoming increasingly clear that most eukaryotic organisms live in close association with different microbial partners. These associations can be neutral (commensalism), beneficial to one partner while detrimental to the other (parasitism) or beneficial for both partners (symbiosis). Many symbiosis systems studied to date have a metabolic function whereby the symbiont provides nutrients the host is unable to synthesise or digest (e.g. pea aphid-*Buchnera,* Baumann *et al.*, 1995; termite gut microbiota, Tholen *et al.*, 2006) or they are able to convert chemical compounds into organic carbon, providing nutrients for their host in habitats in which food is very limited (e.g. hydothermal vent macrofauna, Dubilier *et al.,* 2008). During the evolution of animals, bacteria transitioned from being simply a food source to actually inhabiting body cavities that would evolve into a digestive system. Insects, one of the most diverse animal groups on the planet, are examples of well-established and stable digestive symbioses. These associations are likely to have helped insects colonise such a great variety of habitats and enabled different forms of digestive systems and feeding types to evolve (McFall-Ngai *et al*., 2013). Additionally, they are great model systems for experimental work that can be applied to other systems (Engel and Moran, 2013). The variety of interactions between bacteria and the animal host digestive system, as well as interactions within the bacterial community, can help elucidate molecular signalling pathways and immunity control, which could in turn help, for example, in disease therapy and prevention (Lemaitre and Hoffmann, 2007; Boon *et al.,* 2014). Parallels can be drawn between insects and the most important zooplankton group, copepods, which are equally diverse and widespread (Bron *et al.,* 2011).

The taxonomic composition of gut microbes can vary greatly with symbionts, including microorganisms such as protist, fungi as well as archea and bacteria. The species diversity is usually highest in organisms with more developed guts and specific diets, such as wood-feeding insects (Colman *et al.,* 2012). The community structure can also vary between locations and functions in the gut (Engel and Moran, 2013). The symbiont biomass to host

mass ratio is usually higher in organisms with complex gut systems, whereas animals with simple, long narrow guts usually harbour less diverse communities of symbionts (Colman *et al.*, 2012). Many other factors influence the establishment and community structure of a gut microbiota. The gut is not the most hospitable environment, with a usually varying pH, digestive enzymes, low oxygen concentrations and the host's immune system all contributing to unfavourable conditions for microbial colonisation (Zilber-Rosenberg and Rosenberg, 2008). However, as seen in many animal cases, microbes somehow overcome those obstacles and establish very successful and permanent associations with their hosts. Although hostile in many respects, the gut can also be a safe haven to microbes, especially in terrestrial, dry habitats. This is less important in an aquatic environment where the gut can be viewed as a continuation of the aquatic medium (Hariss, 1993), although it can also be seen as a hotspot of organic matter in an otherwise dilute environment (Tang *et al.*, 2010). If bacteria can overcome the host's innate immunity and establish a successful symbiosis then the gut provides a nutrient rich environment for the symbionts. In many cases, the hosts seem to select a specific bacterial community to inhabit their gut, as has been shown, for example, in stinkbugs (Kikuchi *et al*., 2007), honeybees (Martinson *et al*., 2012) and *Hydra* (Franzenburg *et al*., 2013). However, the most important factors that seem to govern the structure, abundance and role of a stable gut symbiont community are the diet and gut morphology of the host (Ley *et al*., 2008; Perez-Cobas *et al*., 2015).

Compared to the field of insect nutritional symbiosis, little is known about the copepod-associated bacteria. As discussed in Chapter 5, the research of specific copepod-bacteria interactions is fairly limited and has focused mostly on coastal, freshwater and laboratory reared animals. Studies using electron microscopy have observed bacteria attached to the gut lining of various coastal copepods, such as *Eucalanus bungii, Paracalanus* sp. and *Calanus* sp. All of these species had bacterial populations associated with their gut lining or lumen and fecal pellets, although due to the limitation of the methodology, the authors could not determine their function or taxonomy (Nagasawa and Nemoto, 1988; Hansen and Bech, 1996). Other studies have focused on bacterial colonisation of copepod fecal pellets and the effect of food source on the internal bacterial community, showing that the persistence of the gut microbiota relies on a constant supply of bacteria through food intake (De Troch *et al.*, 2010). Similarly, it was found in the copepod *Acartia tonsa* that an external source of bacteria delivered via food particles was necessary to maintain a diverse internal bacterial community (Tang *et al.*, 2009). They also determined that copepod life history, as well as spatial and temporal variations in the environment affecting the food composition had an effect on the copepod-bacteria association.

The aim of this study was to identify and visualise the gut microbiota of *Pleuromamma* copepods using fluorescence *in situ* hybridisation (FISH) in combination with histological sectioning for determining the phylogenetic affiliation of the gut microbiota and an electron microscopy (EM) approach for high resolution images of the bacteria. Several taxa were identified in Chapter 5 as potentially interesting. Among them, the most dominant was a single betaproteobacterial genus, *Limnobacter* and some gammaproteobacterial genera. The hypothesis was that the gut microbiota would be taxonomically diverse; however *Limnobacter* (Betaproteobacteria) was expected to be highly abundant in the gut. It was also considered the most interesting taxon to investigate due to its low abundance in the water-column bacterial community.

6.2 Methods

6.2.1 Sampling

Samples for FISH and EM were collected and preserved as described in Chapter 2. Two individuals (both female) from two stations in the SAG region, stations ZPT-022 and ZPT-024A (see Table 2.2, Chapter 2) were used in this study. The phylogenetic identity of the animals was determined using both morphological characteristics and COI mitochondrial marker genes, as described in Chapter 3.

6.2.2 Limnobacter phylogeny

Nodes related to *Limnobacter* from AMT22 (Appendix C) and AMT24 (Chapter 5) sequencing efforts contained the highest number of sequences in both datasets. Additionally, a 16S rRNA clone library prepared from a female *P. gracilis* DNA extract (AMT22) contained several *Limnobacter*-related sequences. Phylogenetic analysis of the sequences related to *Limnobacter* was done using Geneious v.8.1.7 software (Kearse *et al.*, 2012). Several reference sequences that included free-living *Limnobacter thioxidans*, *L. litoralis* and *L.* sp. (all accession numbers shown in Figure 6.4) and diatom-associated *Limnobacter* sequences (Amin *et al.*, 2015) were downloaded from the NCBI nucleotide database. Other sequences from the class Betaproteobacteria were used in the analysis. These included several animal-related sequences, such as coral and sponge associated Betaproteobacteria (Croue *et al.*, 2013), *Burkholderia* gut symbionts of broad-headed bugs (Kikuchi *et al.*, 2005) and *Daphnia magna* related sequences (Freese and Schink, 2011).

All sequences were aligned using the SINA web-aligner (Pruesse *et al*., 2012) and then truncated to 400 bp (length of Illumina sequences), covering the variable V3-V4 region of 16S rRNA. A *Vibrio atlanticus* sequence downloaded from NCBI was used as an outgroup. Trees were built using Bayesian phylogenetic reconstruction with MrBayes v.3.2.2 (Huelsenbeck and Ronquist, 2001). A General Timer Reversible model with Gamma-distributed rates of evolutions was used with 1,100,000 generations using four parallel Monte Carlo Markov chains and tree subsampling every 200 generations.

6.2.3 Electron microscopy

Transmission electron microscopy (TEM) was performed by Nikolaus Leisch (Symbiosis Department, MPI-MM) using methods described in Chapter 3.

6.2.4 Histology and FISH

Animals fixed in PFA and stored in 50% ethanol were embedded in LR White resin (London Resin Company) in preparation for FISH analysis. Briefly, samples were dehydrated through an increasing ethanol/PBS series of 10 min at each concentration, starting with 60% and finishing with three times 30 min at 100% ethanol. Then samples were put through five gradual resin infiltration steps with the resin concentration increasing from 3:1 to 1:3 ethanol/resin, for 30 min each. The resin infiltration was finished with three 30-minute steps in 100% LR White resin and a last overnight step. Samples were transferred into gelatine capsules and sealed, ensuring no air was trapped inside. Capsules were then placed in a desiccator with approximately 1000 mBar vacuum, flushed with nitrogen gas (to remove all oxygen that prevents the resin from solidifying) and left to polymerise at 50 °C for three days. Embedded samples were then sectioned into 1-µm sections using a Leica EM UC7 ultramicrotome, transferred to Superfrost Plus (Fisher) glass slides and left to dry at 50 °C. Sectioning was done by Oliver Jäckle (MPI-MM).

Sections were hybridised using the general Bacteria (EUB338 I-III, Amann *et al.*, 1990) and specific Betaproteobacteria (Bet42a; together with an unlabelled Gammaproteobacteria competitor, Manz *et al.*, 1992) probes labelled with Cy3 fluorescence dyes (Biomers, Germany). Hybridisations with a Cy5-labelled Gammaproteobacteria probe (Gam42a, with unlabelled Bet42a competitor) were also attempted with limited success. The NONEUB338 probe (Wallner *et al.*, 1993) was used as a negative control. The hybridisation buffer contained 0.9 M NaCl, 20 mM Tris/HCl (pH 8),

35% formamide, 0.01% SDS and probes at an end concentration of 5 ng μl^{-1}. Sections were hybridised for 3 h at 46 °C and then washed for 15 min at 48 °C in a washing buffer containing 70 mM NaCl, 20 mM Tris/HCl (pH 8), 5 mM EDTA and 0.01% SDS. After the washing step, the sections were rinsed with miliQ water and left to air dry. They were then stained with DAPI for 3 min, washed twice in miliQ water for 1 minute and dipped in 96% EtOH, and then left to air dry. Slides were mounted with a Vectashield and Citifluor mixture. They were then examined using an epifluorescence microscope (Nikon, Germany) and Zeiss laser scanning confocal microscope (Zeiss, Germany).

6.2.5 Supplementary information – Appendix C

Additional FISH images showing hybridisations with EUB probe.

A preliminary study, to determine potentially interesting bacteria to target in the gut microbiota analysis, was done on samples collected along a similar Atlantic Meridional Transect in 2010 (AMT22). One *P. gracilis* individual and 27 *P. gracilis/piseki* and *P. borealis* individuals were analysed for associated-bacteria using a clone library and Illumina sequencing, respectively.

Development of a *Limnobacter*-specific FISH probe and images of the formamide test of the Lim836 probe and EUB probes on a pure culture of *Limnobacter thiooxidans*.

6.3 Results

6.3.1 Phylogenetic affiliation of *Limnobacter*-related sequences

The most abundant sequence from two separate Illumina sequencing efforts (MED nodes) and two different cruises (AMT22 and AMT24) was related to *Limnobacter*. Additionally, *Limnobacter*-related sequences were found in a *P. gracilis* derived 16S rRNA clone library. These data are shown in detail in Chapter 5 (AMT24) and Appendix C (AMT22). The phylogenetic relationship of these sequences, when compared to *Limnobacter*-related reference sequences, revealed that both MED nodes and one clone library sequence were basal to the *Limnobacter* clade of the phylogenetic tree of the 400 bp sequences (V3-V4 region of the 16S rRNA) shown in Figure 6.1. The two MED nodes (AMT22_node4639 and AMT24_node12892) and the clone library sequence (AMT22_261) were 100% identical and 99.7% identical to the genus type species *Limnobacter thiooxidans* (accession number NR_025421). Three other clone library sequences were

more variable within the 400 bp region and branched away from the main group of basal sequences (Table 6.1).

Table 6.1: Table showing the percent identity of alignment of *Pleuromamma*-associated clone library (clone) and Illumina sequences (tag node) from AMT22 and AMT24 research cruises and two *Limnobacter* species reference sequences obtained from NCBI GenBank database (accession numbers indicated).

	AMT22 clone 240	AMT22 clone 245	AMT22 clone 261	AMT22 clone 286	AMT22 tag node 671	AMT24 tag node 12892	L. litoralis (AB366172)
AMT22 clone 245	95.9						
AMT22 clone 261	95.8	97.1					
AMT22 clone 286	93.1	93.7	95.9				
AMT22 tag node 671	95.8	97.1	100	95.9			
AMT24 tag node 12892	95.8	97.1	100	95.9	100		
L. litoralis (AB366172)	90.1	91.4	93.9	91.4	93.9	93.9	
L. thiooxidans (AJ289885)	95.5	96.8	99.7	95.5	99.7	99.7	94.2

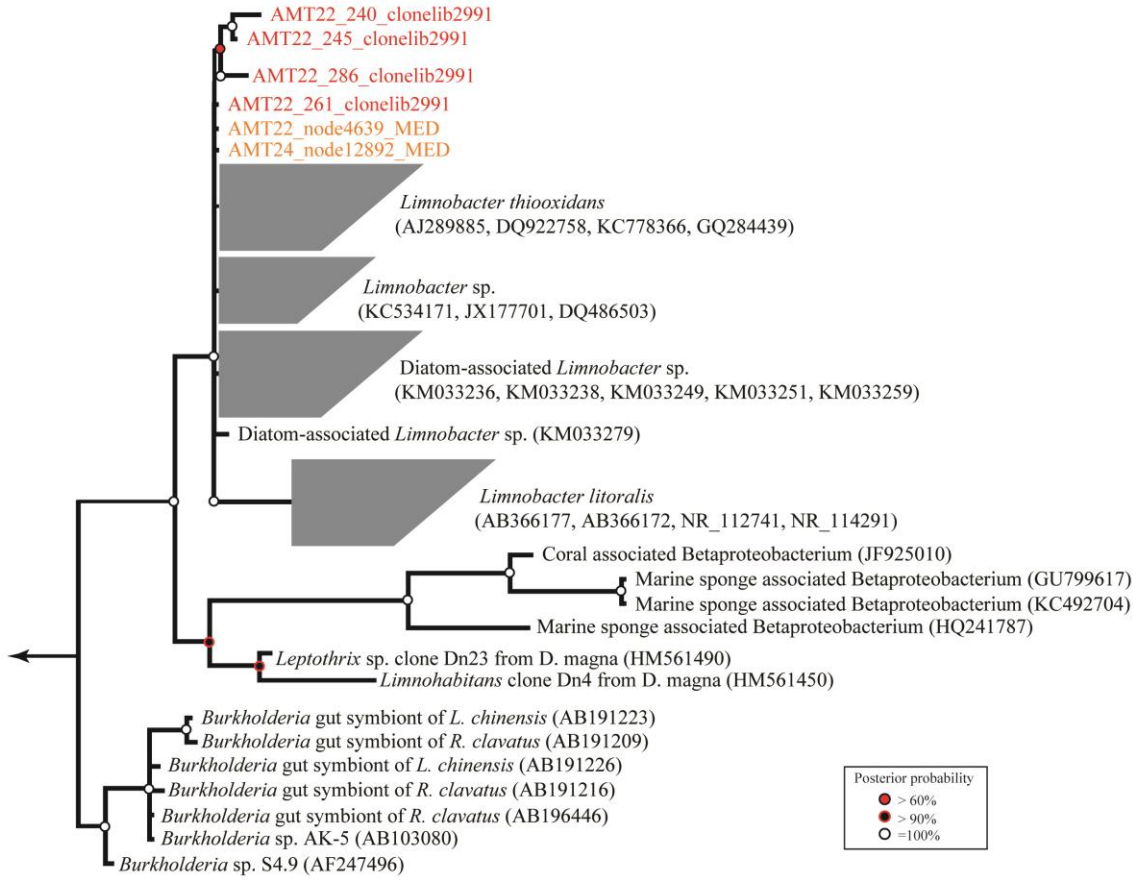


Figure 6.1 Phylogenetic relationship of *Pleuromamma*-associated *Limnobacter* sp. sequences (in colour) to other *Limnobacter* species and animal-associated Betaproteobacteria 16S rRNA sequences obtained from the NCBI GenBank database. AMT22 clone library sequences are shown in red and Illumina sequences in orange. The tree was built using sequences aligned with SINA aligner and Bayesian phylogenetic reconstruction (GTR substitution model) with *Vibrio atlanticus* (Gammaproteobacteria) as an outgroup. Prior to tree calculation, aligned sequences were truncated to 400 bp (length of Illumina sequences). Numbers in brackets are the NCBI accession numbers.

6.3.2 Visualisation of bacteria in the gut of *P. gracilis*

The phylogenetic affiliation of both female individuals presented here was identified as *P. gracilis*. As described in Chapter 3, the gut of the studied individuals was thought to consist of three main regions: foregut, midgut (with three distinct areas) and hindgut. A distinct spatial organisation of bacterial presence was found in the guts of studied *P. gracilis* individuals. The foregut, with a cuticle lining the inner epithelium, was found void of both food particles and bacteria in the sections of *P. gracilis* individuals examined by both TEM and FISH (images not shown).

Visualisation of the *P. gracilis* gut using TEM revealed a large abundance of bacteria in area II of the midgut region (Figure 6.2). The bacteria were present primarily in the lumen and folds of the central, wider area of the midgut (Figure 6.2A). In this area, the surface area of the gut was enlarged by invaginations of the tissue lining, making it distinctly different from the anterior (area I) and posterior (area II) parts of the midgut (Chapter 3). Bacteria observed in TEM images looked viable, with visible, intact membrane structures and some cells in the process of dividing (Figure 6.2B and C). Several different morphotypes of bacterial cells were observed (Figure 6.2C).

The phylogenetic identity of bacteria in the gut of the female *P. gracilis* individual was investigated using a mixture of general bacterial probes (EUB I-III) and a Betaproteobacteria specific probe (Bet42a). Both hybridisations revealed the presence of bacterial signals within the gut, more specifically in area II of the midgut region of the gut (Figure 6.3, 6.4 and Appendix C – Figure C.1). The hybridisations using EUB and Bet42a were done on consecutive sections that were approximately 2-4 µm apart. The images shown in Figure 6.3 are therefore most likely of the same bacterial cells. Hybridisation was also attempted using a specific Gammaproteobacteria probe (Gam42a) but these were mostly unsuccessful with very weak or no signals seen.

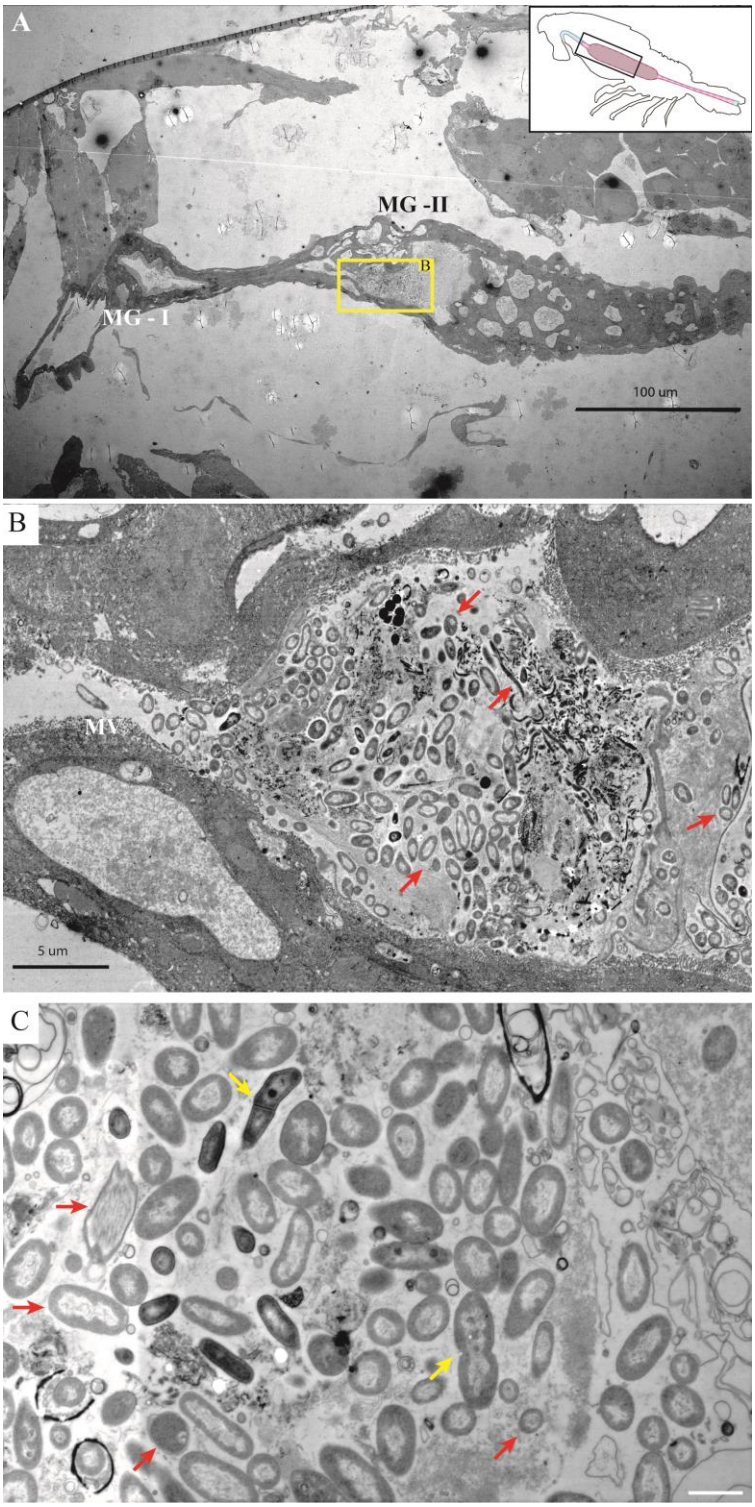


Figure 6.2 Transmission electron micrograph of the midgut area II of a female *P. gracilis* individual. **A**: Overview of the midgut with the anterior midgut (MG-I) and middle midgut (MG-II), including invaginations of the gut tissue. Schematic inlay shows position of images in the copepod. The yellow box highlights enlarged area. **B**: Midgut region with bacterial cells (red arrows). **C:** Close up bacterial cells. Yellow arrows show dividing cells and red arrows show different cell morphotypes. Scale 1 µm.

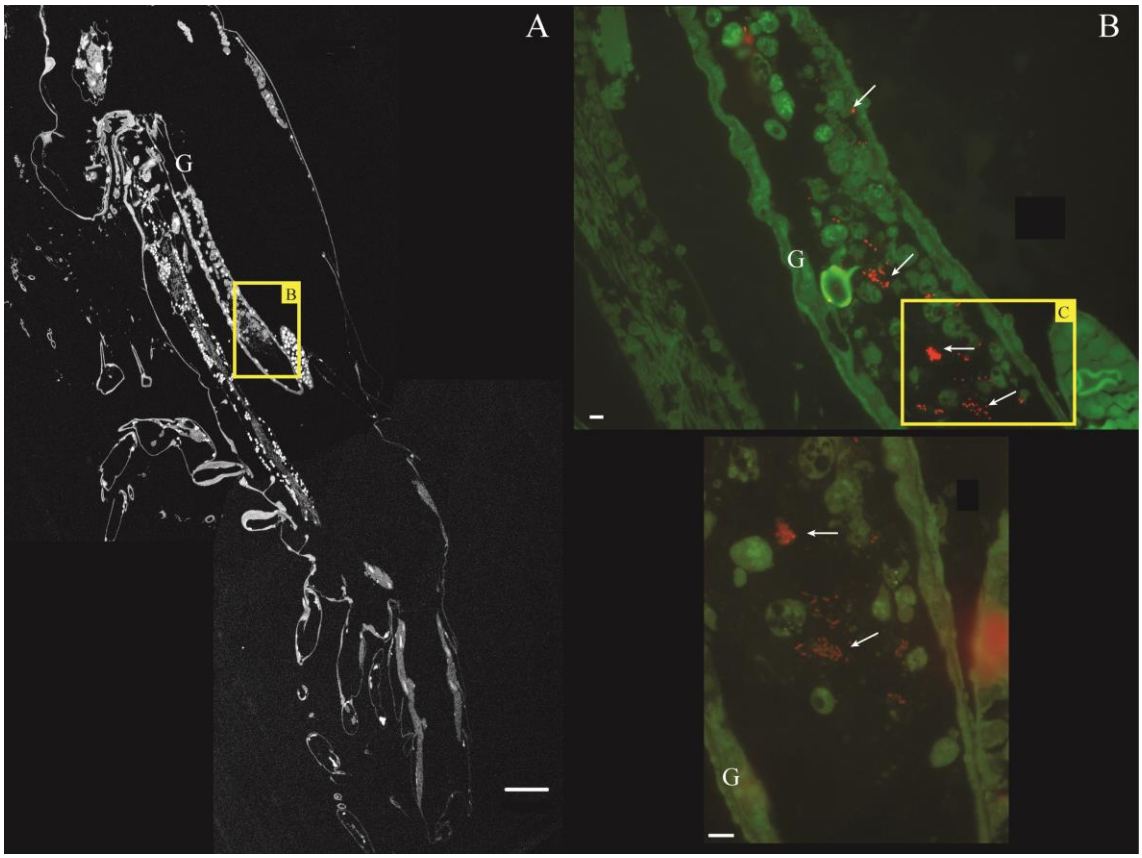


Figure 6.3 EUB and Bet42a FISH on *P. gracilis* LR White section. **A**: Overview image of section at 10x magnification. Host cell nuclei are visible as well as gut outline (G). Yellow box shows enlarged area in B. Scale 100 µm. **B**: Middle midgut at 40x magnification with bacterial cells (white arrows) visible in red using a general bacterial probe (EUBI-III-Cy3). Yellow box is the equivalent area of a different section shown in C. Scale 5 µm. **C**: Middle midgut at 100x magnification. Bacterial cells (white arrows) visible in red were hybridised using the betaproteobacterial probe (Bet42a-Cy3 with an unlabelled Gam42a). Scale 5 µm. The green in panels B and C shows tissue autofluorescence using a FITC filter.

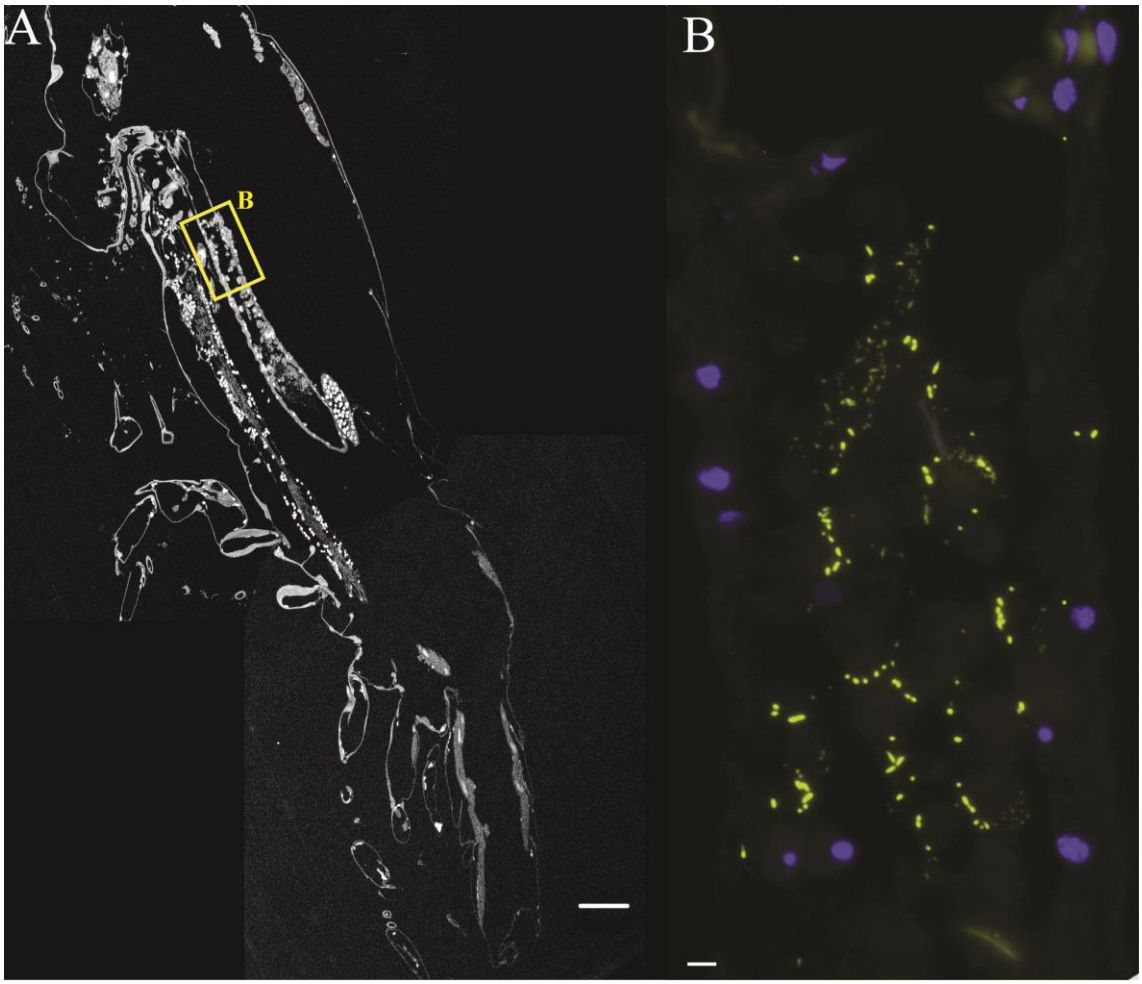


Figure 6.4 EUB FISH on *P. gracilis* LR White section. **A**: Overview image of section at 10x magnification. Host cell nuclei are visible as well as gut outline. Yellow box shows enlarged area in B. Scale 100 µm. **B**: Middle midgut at 40x magnification with bacterial cells visible in yellow using a general bacterial probe (EUBI-III-Cy3). Scale 5 µm. The blue in panel B are DAPI stained host nuclei.

Food particles were visible in the TEM images but no bacterial cells could be detected in the posterior midgut area III (Figure 6.5). Very few bacterial signals were detected with FISH in the section examined and they seemed to be associated with food particles (Figure 6.6).

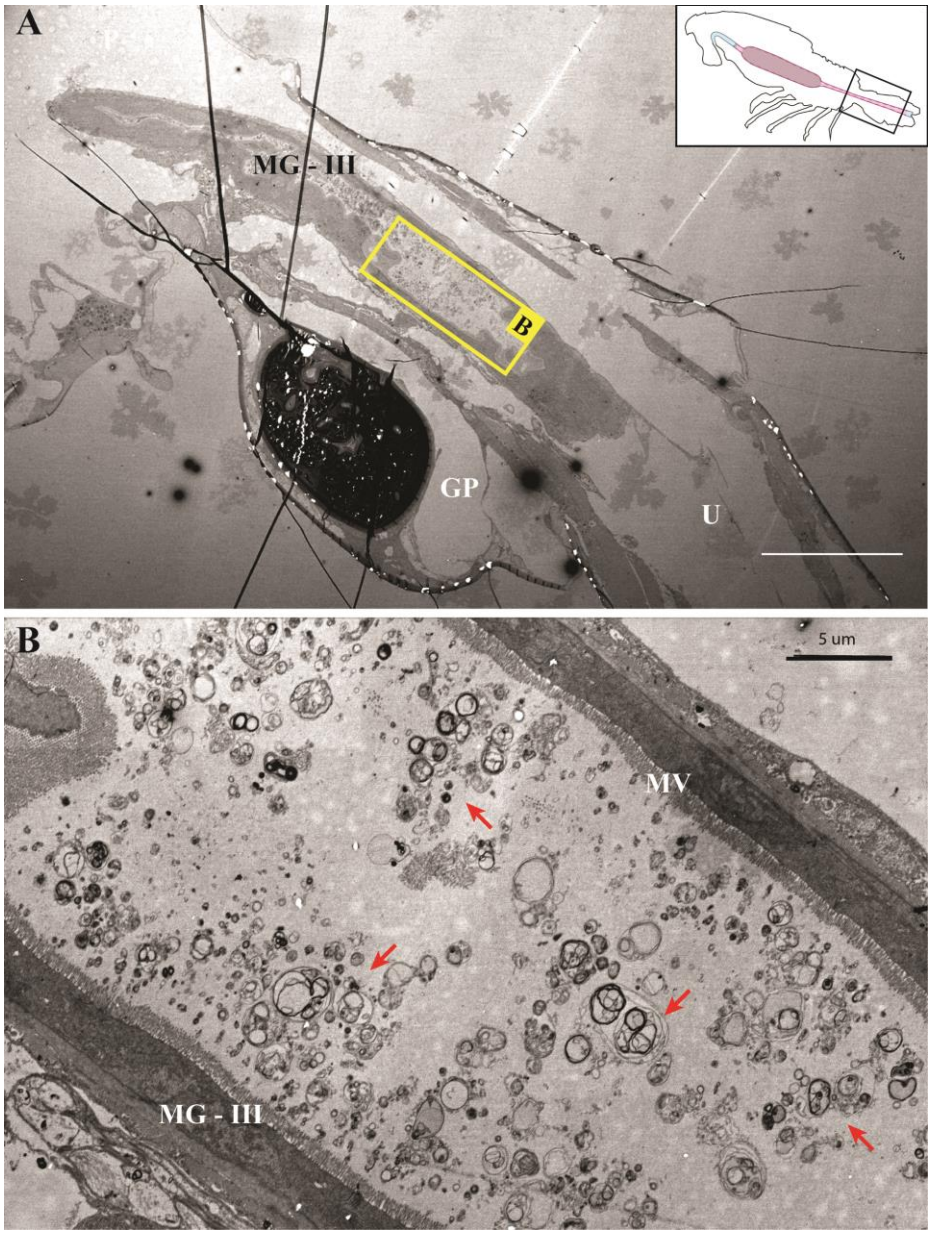


Figure 6.5 Transmission electron micrograph of the midgut area III of a *P. gracilis* female. **A**: Overview image of the posterior midgut (MG-III) visible in the urosome (U). Genital pore (GP). Schematic inlay shows position of images in the copepod. Yellow box highlights the area shown in panel B. Scale 20 µm. **B**: Close up of MG-III with food particles (red arrows) but no visible bacterial cells. The gut is lined with epithelial cells with microvilli (MV). Scale 5 µm.

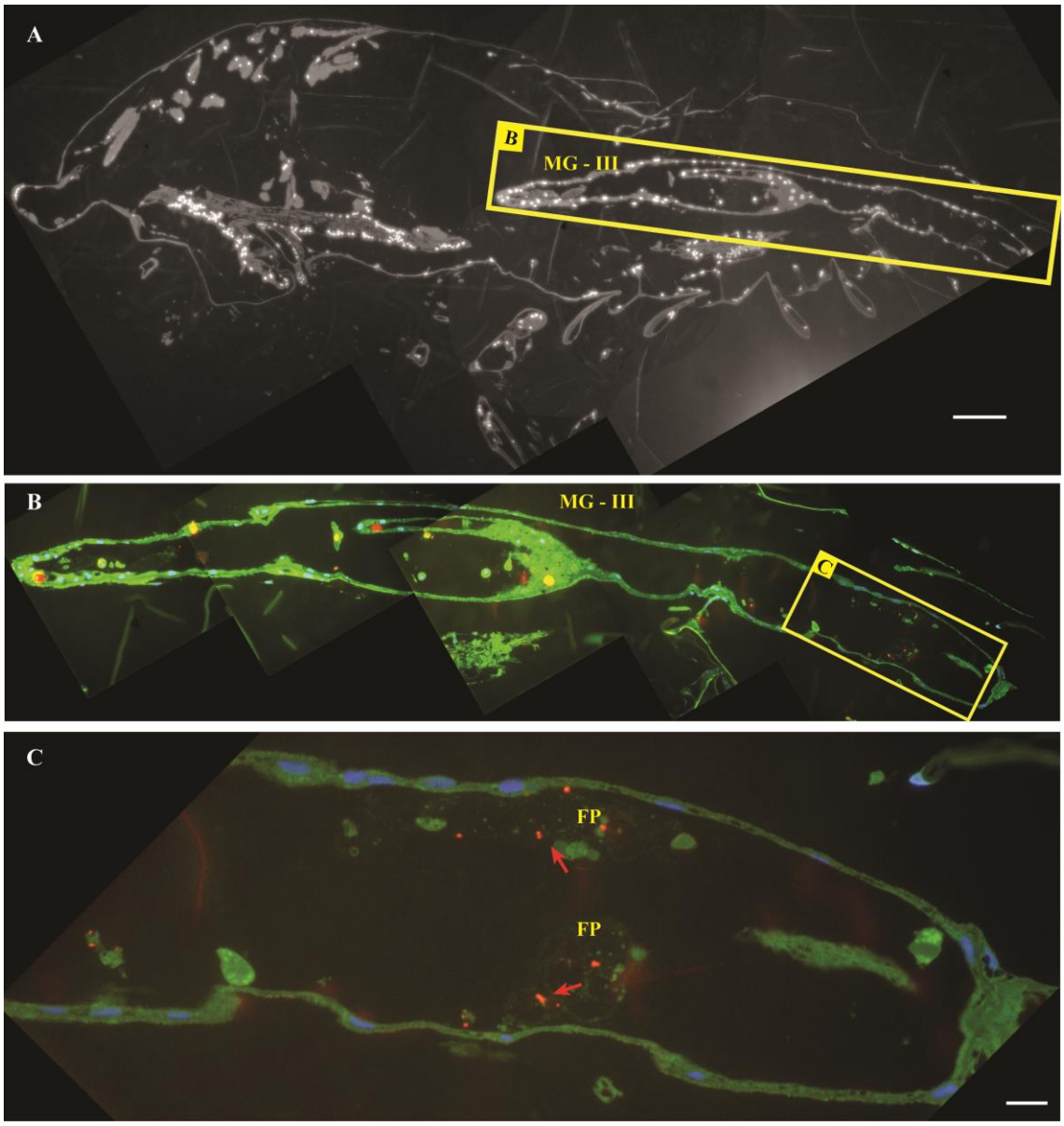


Figure 6.6 FISH on LR White section using EUB338 probe showing the posterior gut of a female *P. gracilis* copepod. **A:** Overview image of the section with DAPI stained host nuclei visible. Yellow box shows area in next panel. Scale 50 µm. **B:** Composite image of the lower gut showing the gut outline (FITC autofluorescence - green) and DAPI stained host nuclei (blue). Bacterial signals can be seen in red (EUB-Cy3) in the posterior part of the gut close to the anal opening attached to food particles (FP). A close up of this area is shown in **C**. Scale 5 µm

6.4 Discussion

The presence of a bacterial community in the gut of *P. gracilis* copepods was identified and partially characterised. This study is the first showing high-resolution electron microscopy images of a bacterial community in the gut of one of the dominant open ocean

copepod genera *Pleuromamma*. Furthermore, no other published study to date has linked sequencing efforts with fluorescent *in situ* hybridisations, thus showing the phylogenetic affiliation of bacterial cells *in situ*.

The highest abundance of bacteria was observed in the middle midgut region, where most digestion is likely to occur (Hasset and Blades-Eckelbarger, 1995). The observation of intact and in some cases dividing bacterial cells led to several possible interpretations: (I) These bacteria have not yet been digested, which is unlikely because of the absence of digested-looking cells; (II) These bacteria could have the capacity to resist digestion and have come from the environment; and (III) These bacteria were part of a resident population possibly interacting with the host, since dividing cells suggest favourable growth conditions. Additionally, several bacterial morphotypes could be observed from the TEM images suggesting various bacterial taxa comprising the gut community. These different taxa could have specific roles within the gut.

The phylogenetic affiliation of the bacterial cells present in the copepod guts was investigated using FISH and hybridised cells in the female *P. gracilis* gut sections were identified as Betaproteobacteria. Two families of cyclopoid copepods were dominated by Betaproteobacteria in a study of North Atlantic copepods (De Corte *et al.*, 2014) and *Sapphirina* sp. from the Sargasso Sea were also shown to be associated with Betaproteobacteria (Shoemaker and Moisander, 2015). However, neither of these studies reported the presence of *Limnobacter*-related sequences. Other examples include the *Crambe crambe* marine sponge, in which a single betaproteobacterial species dominates the sponge tissue, which the authors were able to show using similar sequencing and microscopy-based techniques (Croue *et al.*, 2013). The results of the copepod-associated bacterial sequencing study in Chapter 5 strongly support the hypothesis that the Betaproteobacteria found in this study were most likely *Limnobacter*-related, because this genus was the only highly abundant Betaproteobacteria in the Illumina sequencing dataset. A *Limnobacter*-specific probe has been developed and successfully tested on a pure *L. thiooxidans* culture (Appendix 4). Further experiments on *Pleuromamma* gut tissue will allow confirmation of the taxonomic affiliation of the Betaproteobacteria found in the *P. gracilis* individual of this study.

Additional evidence of the potential importance of *Limnobacter* sp. in the *Pleuromamma* association came from the consistent presence of identical or highly similar sequences (Figure 6.1) across a large spatial scale (the Atlantic Meridional Transect), between two separate sampling years (AMT22 and AMT24) and using different

sequencing methods (clone library and Illumina sequencing). Moreover, very little to no presence of Betaproteobacteria was found in the AMT22 sequencing dataset (Appendix 3) of free-living and particle-attached bacteria along a similar transect to the one sampled in this study. *Limnobacter* sp. has been reported as part of the diatom microbiome (Amin *et al.*, 2015) and this association could represent a vector for colonisation of the copepod gut.

Based on the phylogenetic tree, the *Limnobacter* sequences from this study, were most closely related to the type species of the genus, *L. thiooxidans*. The type strain was isolated from a freshwater lake sediment (Spring *et al.*, 2001) while the second species in the genus, *L. litoralis* was found in a volcanic deposit (Lu *et al.,* 2011). The described species of the *Limnobacter* genus are strictly aerobic, heterotrophic and have the ability to oxidise thiosulfate (Lu *et al.,* 2011). They are also described as slightly curved rods that are motile by a single polar flagellum (Spring *et al.*, 2001). A marine strain or species such as that found in this study could have a completely different ecology.

Further analysis of the *Pleuromamma*-associated *Limnobacter* is needed for a clearer assessment of phylogenetic affiliation (e.g. using high-quality full-length sequences) and determining a metabolic function (e.g. doing metagenomic and metatranscriptomic sequencing). More work (e.g. testing of optimal formamide concentration) is also needed for successful use of the *Limnobacter*-specific probe on tissue sections in order to provide definitive evidence of their presence in the gut of *Pleuromamma* copepods. Due to technical difficulties with the Gammaproteobacteria probe (Gam42a) and limited useful sections, no clear indication was found of gammaproteobacterial presence in the gut bacterial community. A solution that could be tested is the use of whole-mount FISH on an intact gut. The issue with whole-mount FISH on entire copepods is that the chitinous cuticle presents a barrier for the FISH probes and subsequent microscopy. Once the gut is dissected out of the animal, this should not present an issue, especially for the midgut region, which has no cuticle inner lining. Additionally, other general and taxon-specific probes (e.g. Alphaproteobacteria, Bacteriodetes, *Marinobacter*, etc.) could be applied to *Pleuromamma* tissue sections or whole guts to identify other members of the gut bacterial community.

Chapter 7 Overall Conclusions

7.1 Summary

The *in situ* identification and assessment of the spatial distribution of the gut microbiota of an ecologically important genus of oceanic copepods were performed for the first time. The strength and robustness of the study derives from sampling at the basin scale of the Atlantic Ocean, as well as from sampling a high number of individual copepods.

The phylogeny and morphology of the host copepods were investigated to support hypotheses made about the *Pleuromamma* microbiome. The *Pleuromamma* species in this study could be grouped into two distinct clades based on the phylogeny of the mitochondrial COI gene marker. The first was composed of *P. borealis* related sequences and originated mostly from the EQ region. The second was a mixture of *P. gracilis* and *P. piseki* related sequences from all three investigated regions along the AMT cruise track. The phylogenetic structures of the latter two species have previously been reported to form a cryptic species complex, so they were considered as one group for the purpose of this study. Animals sampled in the EQ region were of a similar size to those from the NAG and SAG.

The external morphology of *Pleuromamma* individuals provided information about the potential food source and feeding behaviour. External feeding appendages indicated the ability both to concentrate smaller particles and to crush them. These could range from detrital material, microphytoplankton to microzooplankton, for example diatoms or dinoflagellates. Further evidence of the *Pleuromamma* diet was obtained through Stable Isotope Analysis (SIA) of two microplankton size fractions (20-100 μm and 100-180 μm) and copepods. The pattern of nitrogen stable isotope ($\delta^{15}N$) composition of the microplankton size fractions reflected biological processes occurring in different areas of the Atlantic. Lower $\delta^{15}N$ values were measured north of the Equator, indicating the input and transfer of biologically fixed N_2 through the food web. Higher $\delta^{15}N$ values in the SAG suggested a smaller contribution of nitrogen fixation as the source of nitrogen to higher trophic levels. These patterns were mirrored in the stable isotope composition of *Pleuromamma* copepod samples. Values measured in animals from the NAG and EQ were on average higher than the microplankton baseline, whereas similar values to the baseline were recorded in animals from the SAG. This may suggest that microplankton are more important as a food source for animals in the North Atlantic, while animals in the South

Atlantic rely more on smaller particulate material and therefore do not have a typical higher $\delta^{15}N$ composition than the perceived food source. These results need to be supported by more measurements, though, to make them more robust. Stable isotope measurements of particulate material (0.7-20 μm) would provide additional insight into the diet of the *Pleuromamma* copepods. Although, the diet of the studied copepods could not be determined in detail, some insight has been gained and the data here provide valuable information for further studies. The community composition of the microplankton was also assessed and it revealed that while the estimated abundance may vary among regions along the cruise transect, the community composition remained stable and consisted primarily of *Ceratium* spp., diatoms, *Trichodesmium* (particularly in the North Atlantic), as well as copepod nauplii. Interestingly, the stability in prey community composition was reflected in the stability of *Pleuromamma*-associated gut microbiota.

Illumina tag sequencing was used to identify the bacteria associated with *Pleuromamma* copepods. One of the main hypotheses of this thesis was that the copepod-associated bacterial communities would most likely differ among regions sampled along the AMT transect and between copepod species. This hypothesis could, however, not be confirmed with the data obtained in this study. A comparison of sequence diversity among regions along the cruise transect and between the copepod species proved that indeed there were no statistically significant differences. However, when the community dataset was assigned taxonomic paths, some patterns could be observed when looking at certain taxa. Although a high diversity of sequences was obtained with the sequencing effort, a few dominant taxa were consistently present in high relative abundance. These belonged to the Betaproteobacteria, Alphaproteobacteria, Gammaproteobacteria and Cyanobacteria. The presence of Cyanobacteria and certain gammaproteobacterial genera that are known to associate with detrital material could be explained by *Pleurommama* actively feeding on these taxa. Other Gammaprotobacteria, for example *Vibrionaceae* related sequences, were likely to be externally or internally associated with the copepods. Indeed, *Vibrio* sp. has been described as a common copepod-associated taxon, reported both as epibionts and found in the gut. Other taxa, such as the epsilonproteobacterial genus *Sulfurovum* and deltaproteobacterial genus *Peredibacter*, had distinctly higher relative abundances in both gyre regions. This may indicate that there were some region specific associations but more investigation is needed to reach stronger conclusions.

The copepod-associated bacterial community was compared to free-living and attached water-column bacterial communities sampled along the AMT22 cruise transect. When compared at the bacterial class level the three communities had different profiles with some taxa in common, such as the Gammaproteobacteria and Cyanobacteria. These were likely obtained through copepod feeding. A certain caution should be exercised, though, because the compared datasets were obtained using different sequencing techniques. This analysis showed that certain taxa were indeed shared among the water-column and copepod-associated datasets, however the copepod-associated bacterial community contained unique members that were not found in the surrounding environment. The hypothesis that the water-column and copepod-associate bacterial communities would differ can be partially accepted.

Probably the most interesting sequences observed associated with *Pleuromamma* copepods, but absent from the environment, were those related to Betaproteobacteria, more specifically a single genus *Limnobacter*. Known described species of the genus are freshwater and there are very few published marine sequences related to the genus. These sequences were highly abundant in both the EQ and SAG with a lower relative abundance in the NAG. This may be due to a lower physiological condition of NAG animals compared to equatorial and SAG animals, for example due to the effect of seasonality between the Northern and Southern Hemispheres. Animals in the North were sampled in the boreal autumn when food concentrations may be getting depleted, while animals in the South were sampled during the austral spring. Further evidence of a stable associations comes from the recurrent presence of the taxon in samples from different research cruises (AMT22 and AMT24) as well as detected with different methods (next-generation sequencing and clone library approach).

Investigation into the *in situ* presence of Betaproteobacteria in the gut of *Pleuromamma* copepods confirmed the hypothesis that they did form part of the gut microbiome. This was shown in samples from the SAG. Electron microscopy images also showed a bacterial community comprising of several bacterial morphotypes, so further identification of the phylogenetic identity of these bacteria is needed. The gut microbiome was located in the part of the midgut specialised for digestion and absorption of nutrients. No bacteria could be found in the foregut or the posterior part of the midgut.

7.2 Future outlook

The presented novel and consequently descriptive study has laid a firm foundation by generating hypotheses for future research of copepod-associated bacterial communities and more detailed investigation of the bacterial community structure, as well as analyses of the interactions taking place between all partners. Future work investigating the *Pleuromamma* copepod microbiome could focus on a number of research questions:

(1) What is the nature and specificity of the *Limnobacter*-*Pleuromama* association?
(2) Are there other specific bacteria in the *Pleuromamma* gut?
(3) Is the same *Limnobacter* species or strain consistently present as part of the gut microbiome in different species of *Pleuromamma* and among regions in the Atlantic?

The research questions suggested above constitute a follow up postdoc project. To answer questions about the specificity of the association and the nature of the interactions, a set of twelve samples from both gyres and the equatorial region, and covering all three *Pleuromamma* species have been sent for metatranscriptome and metagenome sequencing. In order to minimise the amplification of the host genome and externally associated bacteria, the guts were dissected from the animals.

Several bacterial morphotypes were identified in the gut and, although Betaproteobacteria were shown to colonise the midgut region, other taxa should be investigated. The most likely candidates to investigate using FISH oligonucleotide probes would be several gammaproteobacterial genera, such as *Marinobacter*, *Vibrio* and *Alteromonas*. Others could include Alpha-, Epsilon- and Delatproteobacteria. A move away from histological sectioning towards using whole-mount FISH on whole copepod gut preparations may also provide better insight into the spatial distribution of the gut microbiome within the entire intestinal tract.

Copepods have been suggested as metabolic hotspots, with the host's active feeding concentrating nutrients in an otherwise oligotrophic environment. Metagenome and metatranscriptome assembly of the gut microbiome will allow the prediction of function and expression of metabolic pathways. Using these data, hypotheses can be formed of the type of association, for example beneficial or pathogenic. Assembly of full-length 16S rRNA genes from the metagenomic libraries will allow more in-depth characterisation of the phylogenetic relationship of key bacterial taxa (e.g., *Limnobacter*) among copepod species

and regions. In many symbiotic associations, symbionts have been found to differ on species or strain level among the same or closely related host species.

(4) Do copepod nauplii and early copepodite stages have a similar gut microbiota?
(5) How is the gut microbiota acquired? What is the transmission mode?

The constant association of the same bacterial taxa in pelagic Atlantic copepods opens up the question of symbiont transmission. In order to fully understand the details of the association, the ultimate tool in symbiosis research is to be able to cultivate both symbionts and host, together as well as separately. This is particularly true for zooplankton research, in which one of the main challenges is that nauplii and copepodite stages are very difficult to identify to species level based on morphology. The best approach would thus be to collect live adult female individuals and attempt to hatch and collect the various life stages, allowing a better control of the research parameters.

The presence of a similar microflora associated with early life stages of the copepods would indicate a specific association between the partners, additionally suggesting a selective advantage to keep the same bacterial partner from one generation to the other. On the other hand, the absence of bacteria would suggest a plastic ability to associate with a bacterial population adapted to the local environment.

Being able to manipulate the association directly would also allow investigation into the benefits the gut microbiome could have on their copepod host. Removing the gut-associated bacteria could indicate whether these bacteria have an impact on the host's fitness. While feeding experiments would show whether the gut microbiome stays stable independently of the environment or whether it changes with it.

7.3 Conclusions

The present work showed the recurrent association between *Pleuromamma* copepods and bacteria in the Atlantic Ocean, and discussed the possibility of some being part of the host's diet, colonising the the host's gut and carapax. Studying interactions in the pelagic trophic web is extremely important in order to be able to predict its response to changing conditions. For example, recent time series have shown that the oligotrophic ocean is exapanding (Polovina *et al.*, 2008) and it is therefore important to understand what the impact may be on organisms inhabiting these regions. Copepods are the most abundant zooplankton group and have a direct affect on organisms at both a lower and higher trophic level. Having an associated microbiota has been shown to positively affect a

number of animals (e.g. aiding in digestions, increasing fecundity, etc.) and it has even been hypothesised that association between Bacteria and Insects is what allowed them to adapt to such a range of habitats (Engel and Moran, 2013; McFall-Ngei *et al*., 2013). Understanding the associations between copepods and bacteria in the pelagic environment could thus help to predict the response of these pivotal organisms to future environmental perturbations. The results of this study have enabled several hypothesis to be generated about the nature of the association and how to proceed with future research.

Appendices

Appendix A

Supplementary material to Chapter 3

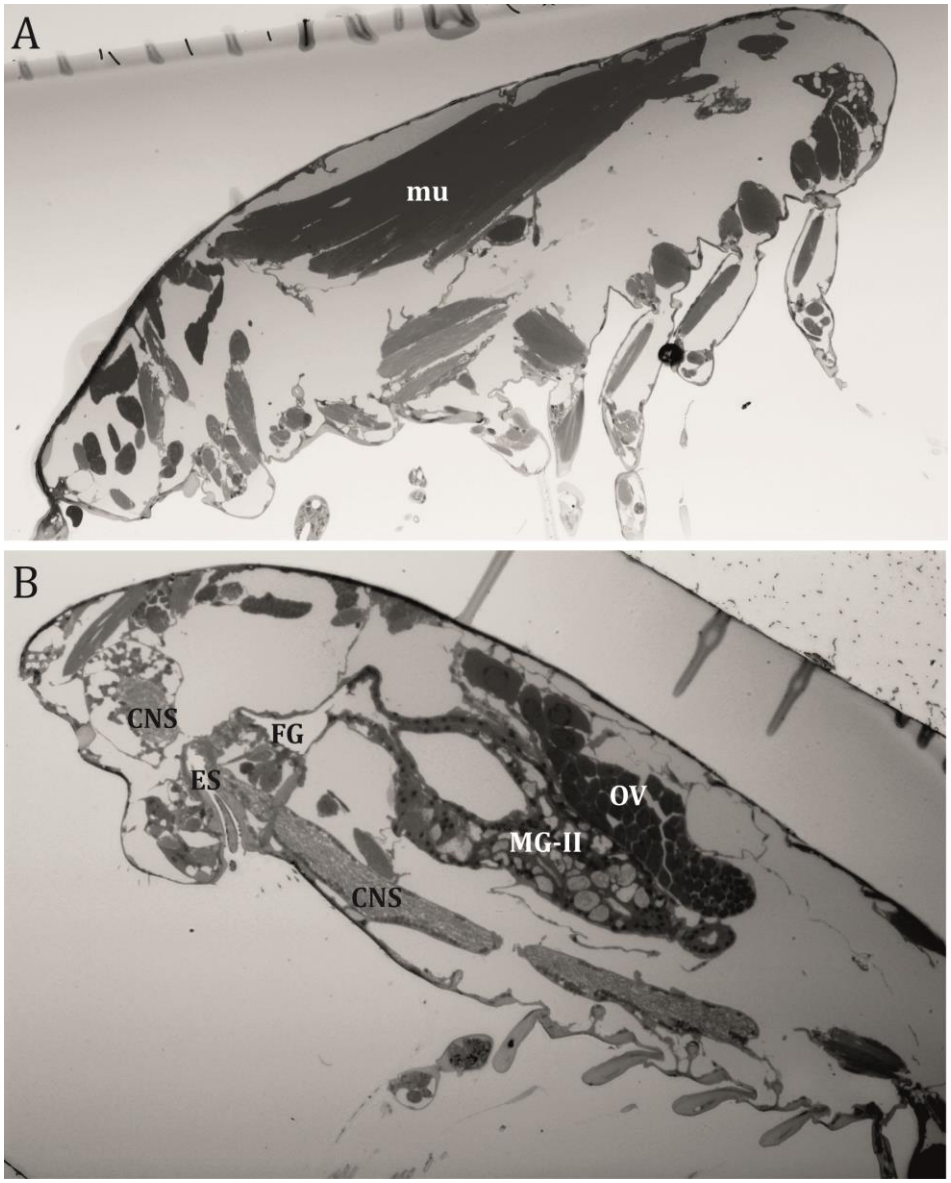


Figure A.1 Additional images of the inner morphology of a *P. gracilis* female copepod. **A:** Section showing muscle tissue. B: Section showing various gut regions, central nervous system and ovaries. **Abbreviations** FG – foregut. MG – midgut. CNS – central nervous system. ES – esophagus. mu – muscle tissue.

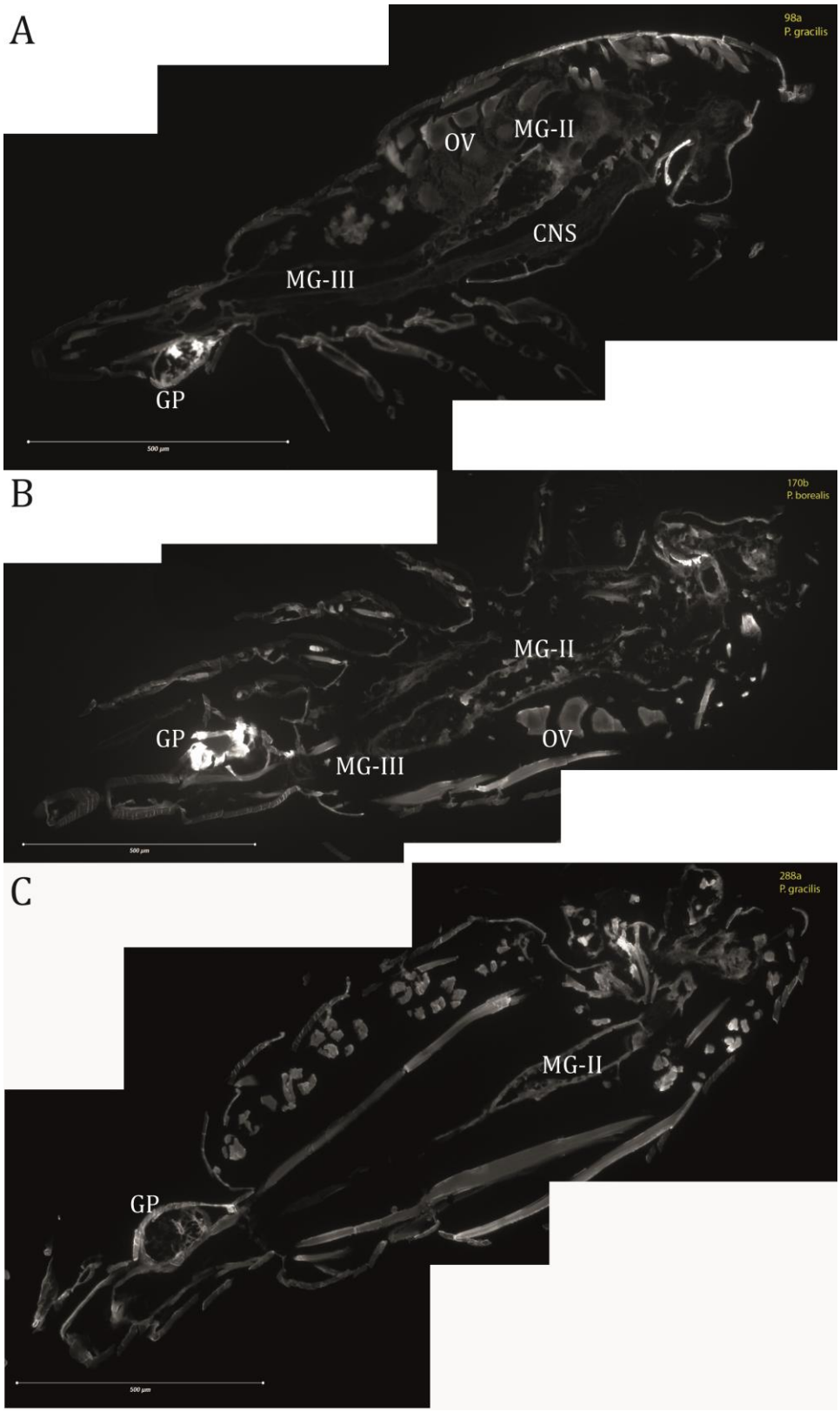


Figure A.2 Paraffin sections of *P. garcilis* and *P.* borealis copepods. **Abbreviations** MG – midgut. GP – genital pore. OV – ovaries. CNS – central nervous system.

Appendix B

Supplementary material to Chapter 5

Appendix B.1 Additional tables and figures complementing AMT24 *Pleuromamma*-associated bacterial community composition results.

Table B.1 Primer, barcode and heterogeneity spacer (HS) sequences for the 27 forward (319F) primers.

Barcode number	Primer	Barcode	HS	Primer sequence
1	319-F	TTATCCG		ACTCCTACGGGAGGCAGCAG
2	319-F	TTATGGC		ACTCCTACGGGAGGCAGCAG
3	319-F	TTACACC		ACTCCTACGGGAGGCAGCAG
4	319-F	TTACGTG		ACTCCTACGGGAGGCAGCAG
5	319-F	TTAGAGG	T	ACTCCTACGGGAGGCAGCAG
6	319-F	TTAGCTC	T	ACTCCTACGGGAGGCAGCAG
7	319-F	TTCTTCC	T	ACTCCTACGGGAGGCAGCAG
8	319-F	TTCTCGT	T	ACTCCTACGGGAGGCAGCAG
9	319-F	TTCTGAG	GT	ACTCCTACGGGAGGCAGCAG
10	319-F	TTCAACG	GT	ACTCCTACGGGAGGCAGCAG
11	319-F	TTCAGTC	GT	ACTCCTACGGGAGGCAGCAG
12	319-F	TTCCTGG	GT	ACTCCTACGGGAGGCAGCAG
13	319-F	TTCCGCT	CGA	ACTCCTACGGGAGGCAGCAG
14	319-F	TTCGCAA	CGA	ACTCCTACGGGAGGCAGCAG
15	319-F	TTGTCAC	CGA	ACTCCTACGGGAGGCAGCAG
16	319-F	TTGATGC	CGA	ACTCCTACGGGAGGCAGCAG
17	319-F	TTGACCT	ATGA	ACTCCTACGGGAGGCAGCAG
18	319-F	TTGCAGA	ATGA	ACTCCTACGGGAGGCAGCAG
19	319-F	TTGGTTG	ATGA	ACTCCTACGGGAGGCAGCAG
20	319-F	TATTCGC	ATGA	ACTCCTACGGGAGGCAGCAG
21	319-F	TATAGCC	TGCGA	ACTCCTACGGGAGGCAGCAG
22	319-F	TATCAGG	TGCGA	ACTCCTACGGGAGGCAGCAG
23	319-F	TATGCTG	TGCGA	ACTCCTACGGGAGGCAGCAG
24	319-F	TAACTCG	TGCGA	ACTCCTACGGGAGGCAGCAG
25	319-F	TAACCAC	GAGTGG	ACTCCTACGGGAGGCAGCAG
26	319-F	TAACGGT	GAGTGG	ACTCCTACGGGAGGCAGCAG
27	319-F	TAAGTGC	GAGTGG	ACTCCTACGGGAGGCAGCAG

Table B.2 Primer, barcode and heterogeneity spacer (HS) sequences for the 27 reverse (806-R) primers.

Barcode number	Primer	Barcode	HS	Primer sequence
1	806-R	TAGCCTA		GGACTACHVGGGTWTCTAAT
2	806-R	TAGGCGT		GGACTACHVGGGTWTCTAAT
3	806-R	TCTTACG		GGACTACHVGGGTWTCTAAT
4	806-R	TCTTGTC		GGACTACHVGGGTWTCTAAT
5	806-R	TCTACCA	A	GGACTACHVGGGTWTCTAAT
6	806-R	TCTAGGT	A	GGACTACHVGGGTWTCTAAT
7	806-R	TCTCTTG	A	GGACTACHVGGGTWTCTAAT
8	806-R	TCTGTGA	A	GGACTACHVGGGTWTCTAAT
9	806-R	TCTGAAC	TC	GGACTACHVGGGTWTCTAAT
10	806-R	TCATCGA	TC	GGACTACHVGGGTWTCTAAT
11	806-R	TCAATCC	TC	GGACTACHVGGGTWTCTAAT
12	806-R	TCACGCA	TC	GGACTACHVGGGTWTCTAAT
13	806-R	TCAGACT	CTA	GGACTACHVGGGTWTCTAAT
14	806-R	TCCTCTG	CTA	GGACTACHVGGGTWTCTAAT
15	806-R	TCCATAG	CTA	GGACTACHVGGGTWTCTAAT
16	806-R	TCCGGTT	CTA	GGACTACHVGGGTWTCTAAT
17	806-R	TCGTTCT	GATA	GGACTACHVGGGTWTCTAAT
18	806-R	TCGAATG	GATA	GGACTACHVGGGTWTCTAAT
19	806-R	TCGAGAA	GATA	GGACTACHVGGGTWTCTAAT
20	806-R	TCGCTAC	GATA	GGACTACHVGGGTWTCTAAT
21	806-R	TGTTCAG	ACTCA	GGACTACHVGGGTWTCTAAT
22	806-R	TGTTGCT	ACTCA	GGACTACHVGGGTWTCTAAT
23	806-R	TGTATCG	ACTCA	GGACTACHVGGGTWTCTAAT
24	806-R	TGTCTGT	ACTCA	GGACTACHVGGGTWTCTAAT
25	806-R	TGTCCTC	TTCTCT	GGACTACHVGGGTWTCTAAT
26	806-R	TGTGACA	TTCTCT	GGACTACHVGGGTWTCTAAT
27	806-R	TGAACTG	TTCTCT	GGACTACHVGGGTWTCTAAT

Table B.3 MED analysis diagnostics with input parameters (-M: Minimum substantive abundance and –V: Maximum variation allowed in each node) and the output (number of outliers represents the sum of sequences excluded due to violation of parameters –M and –V).

Input parameters	
-M	100
-V	4
Number of sequences	1,134,484
Number of samples	79
Number of characters	400
Output	
Total number of outliers	136,350
Number of sequences after removal of outliers	998,134
Number of final nodes	**4,853**

Table B.4 List of samples and explanatory variables (i.e. Station, Region and Copepod species) used in *Pleuromamma* copepod-associated bacterial community composition.

Sample	Station	Region	Latitude	Longitude	Copepod	Sex
cop1	ZPT-006A	NAG	31.8716	-27.55251667	PLGR	F
cop2	ZPT-006A	NAG	31.8716	-27.55251667	PLGR	F
cop3	ZPT-006A	NAG	31.8716	-27.55251667	PLGR	F
cop4	ZPT-006A	NAG	31.8716	-27.55251667	PLGR	F
cop5	ZPT-006A	NAG	31.8716	-27.55251667	PLGR	F
cop6	ZPT-006A	NAG	31.8716	-27.55251667	PLGR	F
cop7	ZPT-006A	NAG	31.8716	-27.55251667	PLGR	F
cop8	ZPT-006A	NAG	31.8716	-27.55251667	PLGR	F
cop9	ZPT-007	NAG	27.04963333	-28.87195	PLGR	F
cop10	ZPT-007	NAG	27.04963333	-28.87195	PLGR	F
cop11	ZPT-007	NAG	27.04963333	-28.87195	PLGR	F
cop12	ZPT-008A	NAG	24.62178333	-29.74623333	PLGR	M
cop13	ZPT-008A	NAG	24.62178333	-29.74623333	PLGR	ND
cop14	ZPT-008A	NAG	24.62178333	-29.74623333	PLGR	ND
cop15	ZPT-008A	NAG	24.62178333	-29.74623333	PLGR	M
cop16	ZPT-008A	NAG	24.62178333	-29.74623333	PLGR	F
cop17	ZPT-011	EQ	14.20993333	-27.92308333	ND	M
cop18	ZPT-011	EQ	14.20993333	-27.92308333	ND	M
cop19	ZPT-011	EQ	14.20993333	-27.92308333	ND	M
cop20	ZPT-011	EQ	14.20993333	-27.92308333	PLGR	F
cop21	ZPT-011	EQ	14.20993333	-27.92308333	PLBO	F
cop22	ZPT-011	EQ	14.20993333	-27.92308333	PLBO	F
cop23	ZPT-011	EQ	14.20993333	-27.92308333	PLBO	F
cop24	ZPT-011	EQ	14.20993333	-27.92308333	ND	M
cop25	ZPT-011	EQ	14.20993333	-27.92308333	PLGR	F
cop26	ZPT-013	EQ	7.28779	-26.49099	ND	F
cop27	ZPT-013	EQ	7.28779	-26.49099	PLGR	F
cop28	ZPT-013	EQ	7.28779	-26.49099	PLGR	ND
cop29	ZPT-013	EQ	7.28779	-26.49099	PLGR	F
cop30	ZPT-013	EQ	7.28779	-26.49099	PLBO	F
cop31	ZPT-013	EQ	7.28779	-26.49099	PLGR	F
cop32	ZPT-013	EQ	7.28779	-26.49099	PLBO	F
cop33	ZPT-013	EQ	7.28779	-26.49099	PLBO	F
cop34	ZPT-013	EQ	7.28779	-26.49099	PLBO	F
cop35	ZPT-015	EQ	0.080333333	-25.877	PLBO	F
cop36	ZPT-015	EQ	0.080333333	-25.877	PLBO	F
cop37	ZPT-015	EQ	0.080333333	-25.877	PLBO	F
cop39	ZPT-015	EQ	0.080333333	-25.877	PLBO	M
cop40	ZPT-015	EQ	0.080333333	-25.877	ND	M
cop41	ZPT-015	EQ	0.080333333	-25.877	ND	M
cop42	ZPT-016	EQ	-3.886966667	-25.02133333	PLBO	F
cop43	ZPT-016	EQ	-3.886966667	-25.02133333	PLGR	F

Sample	Station	Region	Latitude	Longitude	Copepod	Sex
cop44	ZPT-016	EQ	-3.886966667	-25.02133333	ND	ND
cop46	ZPT-016	EQ	-3.886966667	-25.02133333	ND	M
cop47	ZPT-017	EQ	-7.471045	-25.03686167	ND	ND
cop48	ZPT-017	EQ	-7.471045	-25.03686167	ND	ND
cop49	ZPT-017	EQ	-7.471045	-25.03686167	PLBO	ND
cop50	ZPT-017	EQ	-7.471045	-25.03686167	PLBO	F
cop51	ZPT-018A	SAG	-10.43833333	-25.04726667	ND	M
cop52	ZPT-018A	SAG	-10.43833333	-25.04726667	PLBO	M
cop53	ZPT-018A	SAG	-10.43833333	-25.04726667	ND	M
cop54	ZPT-018A	SAG	-10.43833333	-25.04726667	ND	ND
cop55	ZPT-018A	SAG	-10.43833333	-25.04726667	ND	ND
cop56	ZPT-018A	SAG	-10.43833333	-25.04726667	ND	ND
cop57	ZPT-018A	SAG	-10.43833333	-25.04726667	PLBO	F
cop58	ZPT-018A	SAG	-10.43833333	-25.04726667	ND	M
cop59	ZPT-018A	SAG	-10.43833333	-25.04726667	PLBO	F
cop60	ZPT-020	SAG	-18.31643333	-25.08298333	PLGR	M
cop61	ZPT-020	SAG	-18.31643333	-25.08298333	PLGR	M
cop62	ZPT-020	SAG	-18.31643333	-25.08298333	PLGR	ND
cop63	ZPT-020	SAG	-18.31643333	-25.08298333	PLGR	ND
cop64	ZPT-020	SAG	-18.31643333	-25.08298333	PLGR	ND
cop65	ZPT-020	SAG	-18.31643333	-25.08298333	PLGR	M
cop66	ZPT-020	SAG	-18.31643333	-25.08298333	PLGR	F
cop67	ZPT-020	SAG	-18.31643333	-25.08298333	PLGR	M
cop68	ZPT-024	SAG	-30.5915	-25.8048	PLGR	M
cop69	ZPT-024	SAG	-30.5915	-25.8048	PLGR	M
cop70	ZPT-024	SAG	-30.5915	-25.8048	PLGR	M
cop71	ZPT-024	SAG	-30.5915	-25.8048	PLGR	ND
cop72	ZPT-024	SAG	-30.5915	-25.8048	PLGR	ND
cop73	ZPT-024	SAG	-30.5915	-25.8048	PLGR	ND
cop74	ZPT-024	SAG	-30.5915	-25.8048	PLGR	M
cop75	ZPT-024	SAG	-30.5915	-25.8048	PLGR	F
cop76	ZPT-024	SAG	-30.5915	-25.8048	PLGR	F
cop77	ZPT-024	SAG	-30.5915	-25.8048	PLGR	F
cop78	ZPT-024	SAG	-30.5915	-25.8048	PLGR	F
cop79	ZPT-024	SAG	-30.5915	-25.8048	PLGR	M
cop80	ZPT-024	SAG	-30.5915	-25.8048	PLGR	F
cop81	ZPT-024	SAG	-30.5915	-25.8048	PLGR	F

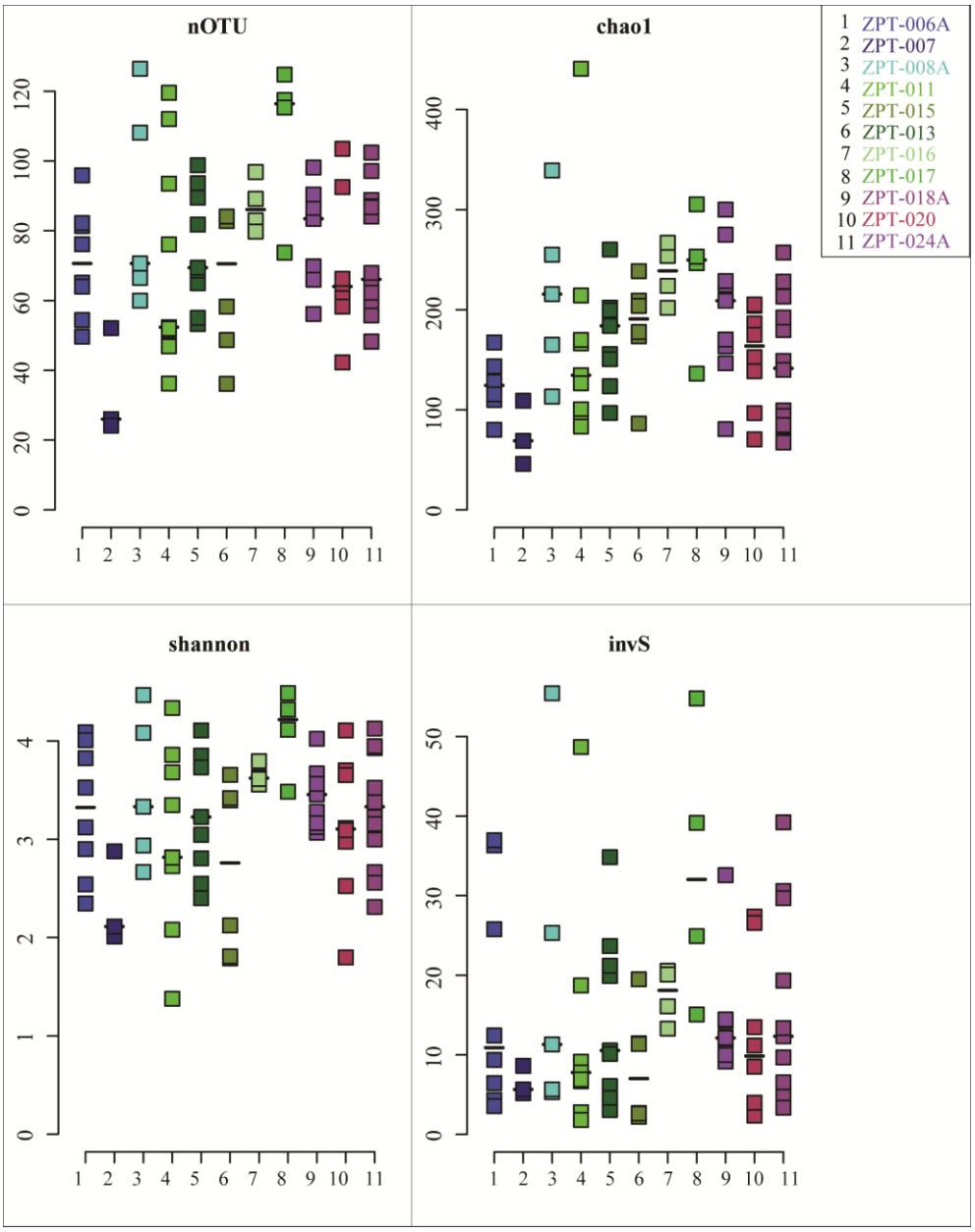


Figure B.1 Alpha diversity of individual samples. Colours show individual stations with blue shades representing the NAG region, green EQ and purple SAG region. nOTU: number of nodes. chao1: Chao 1 diversity index. Shannon: Shannon entropy. invS: Inverse Simpson diversity index.

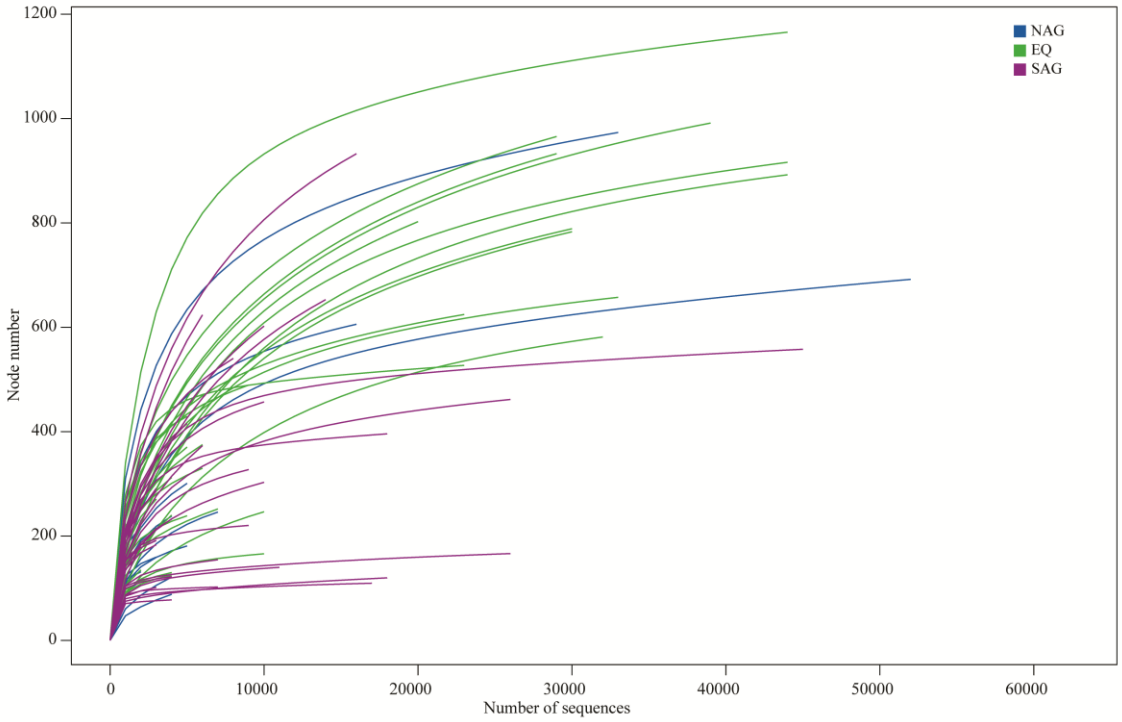


Figure B.2 Rarefaction curves estimating sequencing efficiency. Each curve shows an individual sample, while colours represent regions along the transect

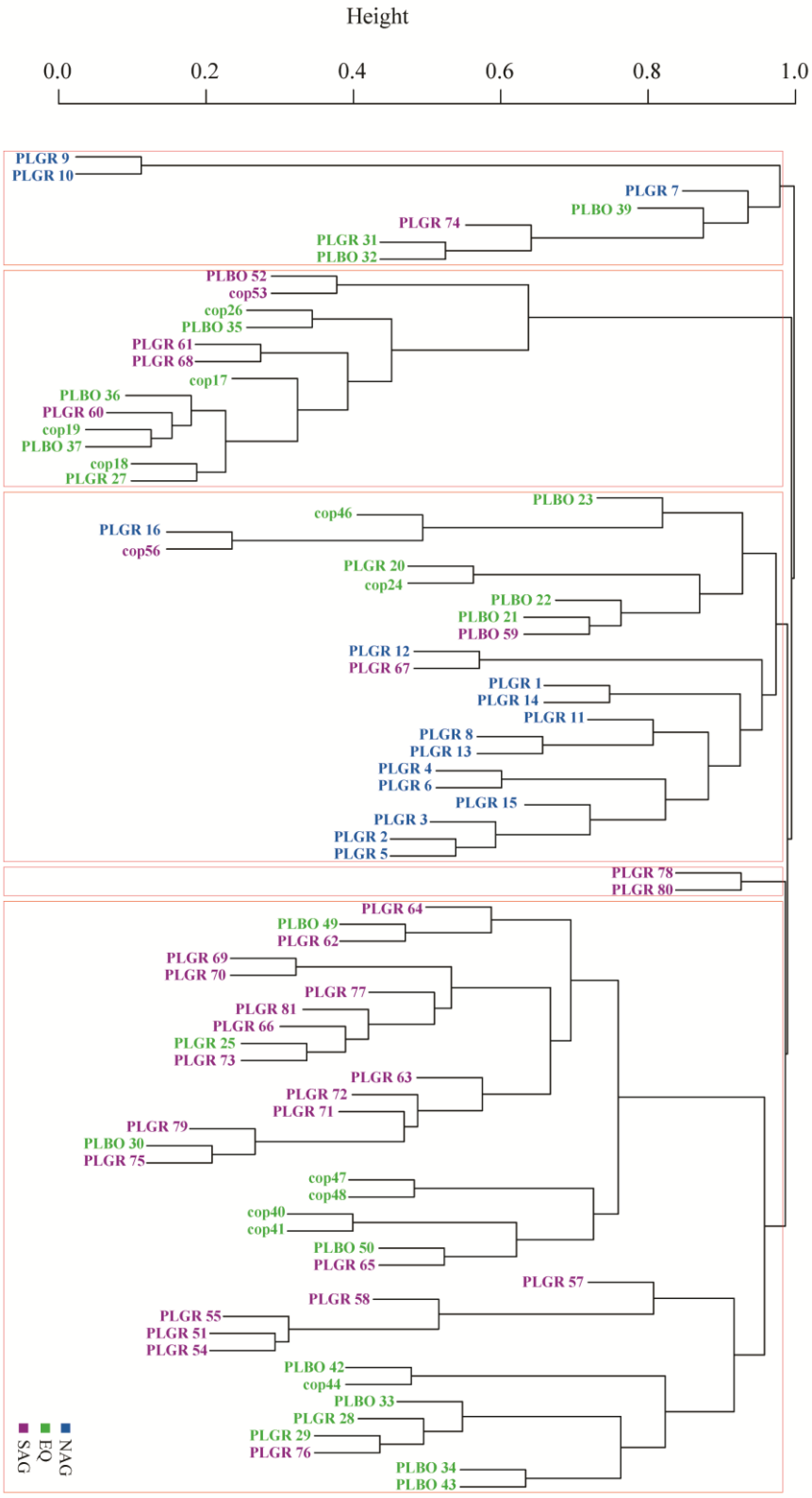


Figure B.3 Cluster dendrogram based on Bray-Curtis dissimilarity matrix of filtered read dataset. Branch labels show copepod species: PLGR – *P. gracilis* and PLBO – *P. borealis*, copXX – species not known. Colours of labels represent regions along the transect. Red boxes show groups at 98% dissimilarity level

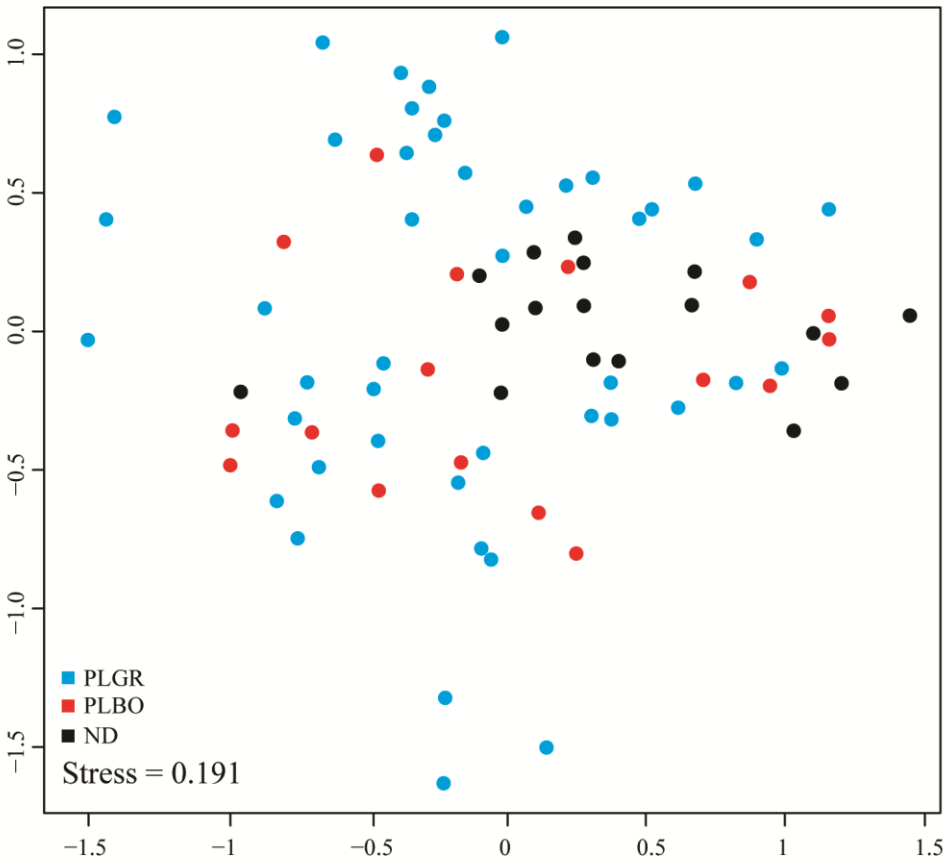


Figure B.4 NDMS plot (using Bray-Curtis method) with data points coloured by copepod species. PLGR – *P. gracilis* and PLBO – *P. borealis*, ND – species not known. Convex hulls (dashed lines) show groups determined at 98% dissimilarity from the cluster diagram in Figure A3.3. Stress value of the NMDS plot is shown in bottom the left corner.

Appendix B.2 Additional figures descrbing the AMT22 (2012) water-column bacterial community composition. Unpublished data, courtesy of Greta Reintjes.

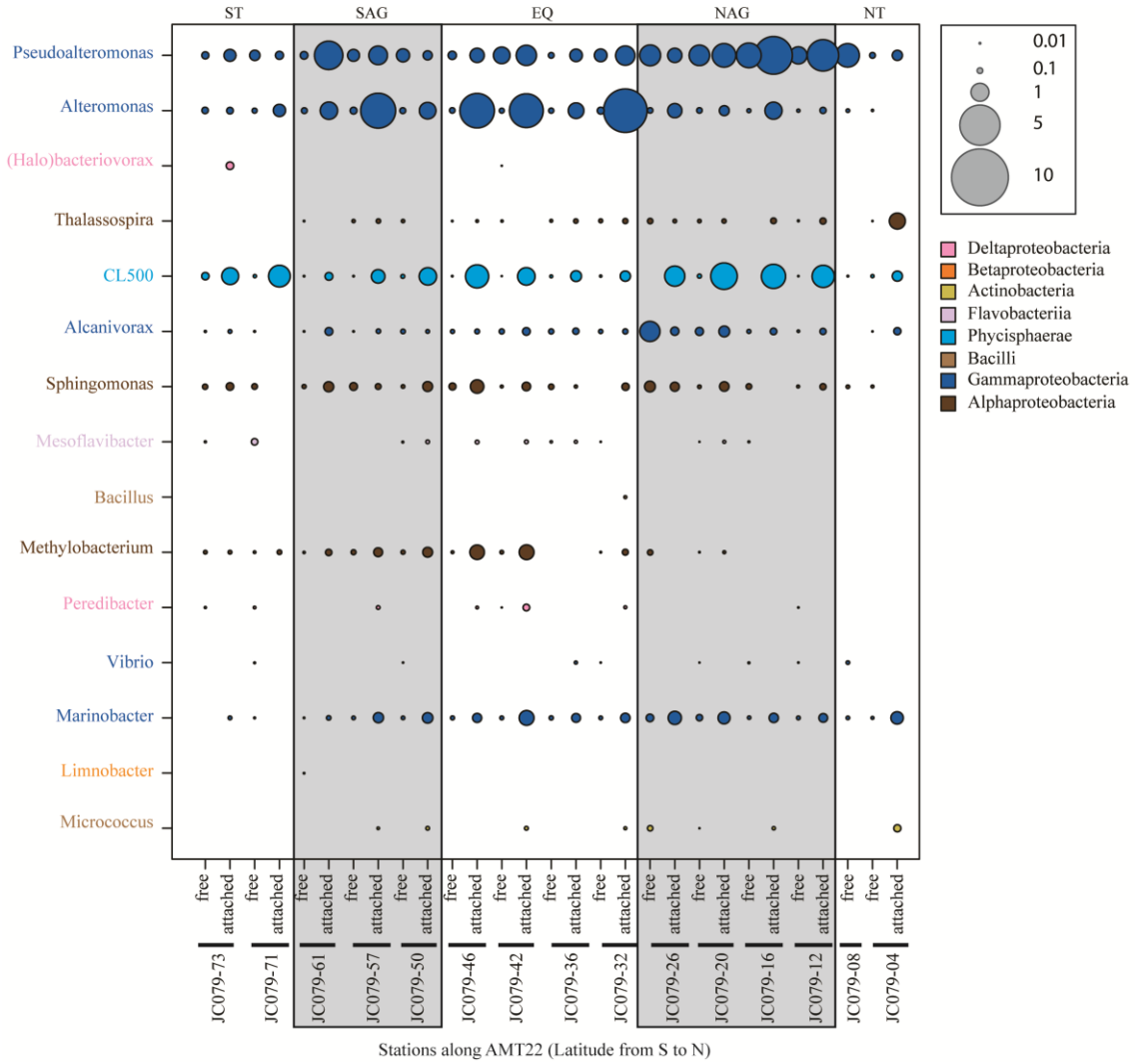


Figure B.5 Water-column bacterial community compositon - relative abundance of genera abundantly present in the copepod-associated bacterial community dataset. Colours correspond to classes and size of circles show relative abundance.

Appendix C

Supplementary material to Chapter 6

Appendix C.1 Additional FISH image

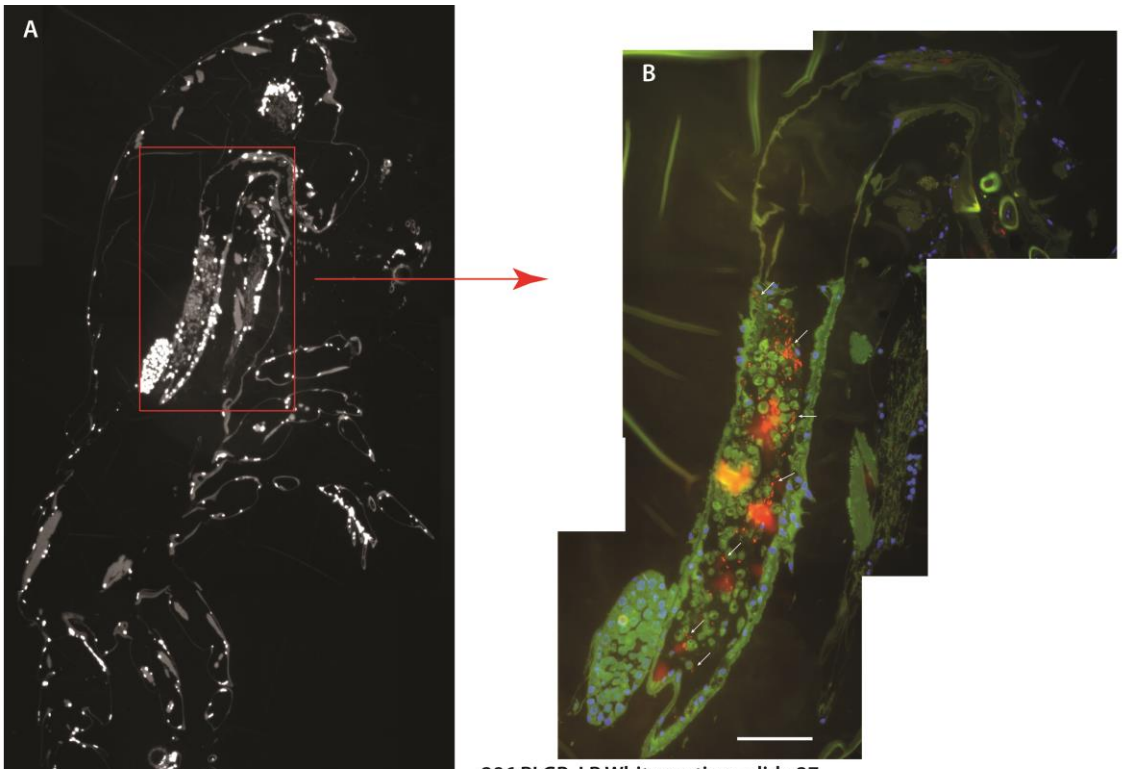


Figure C.1 Additional image showing EUB-Cy3 (red) bacterial signals in the midgut of a female *P. gracilis* copepod.

Appendix C.2 Supporting information for AMT22 Illumina tag sequencing study

A first assessment of the *Pleuromamma* associated bacteria was done using a clone library approach. Briefly, DNA was extracted from a single female *P. gracilis* individual (confirmed by morphology and COI sequencing) following the modified Zhou *et al.* (1996) protocol as described in Chapter 5. General bacterial 16S rRNA primers S-D-Bact-0008-a-S-16/S-D-Bact-1492-a-A-16 (Klindworth *et al.*, 2013) were used for amplification (standard PCR protocol with annealing temperature at 48 °C) and PCR products were cloned using the TOPO TA Cloning Kit (Fisher, Germany). A total of 53 positive clones were sequenced using M13 primer pair and the Big Dye (Fisher, Germany) sequencing kit (as described in Chapter 4). Phylogenetic analysis of sequences was done using the Geneious

v.8.1.7 software (Kearse *et al.*, 2012) and taxonomic affiliation determined using the SINA web-aligner tool (Pruesse *et al.,* 2012). Tree building was done as described in Chapter 6 and result are shown in Figure C.2.

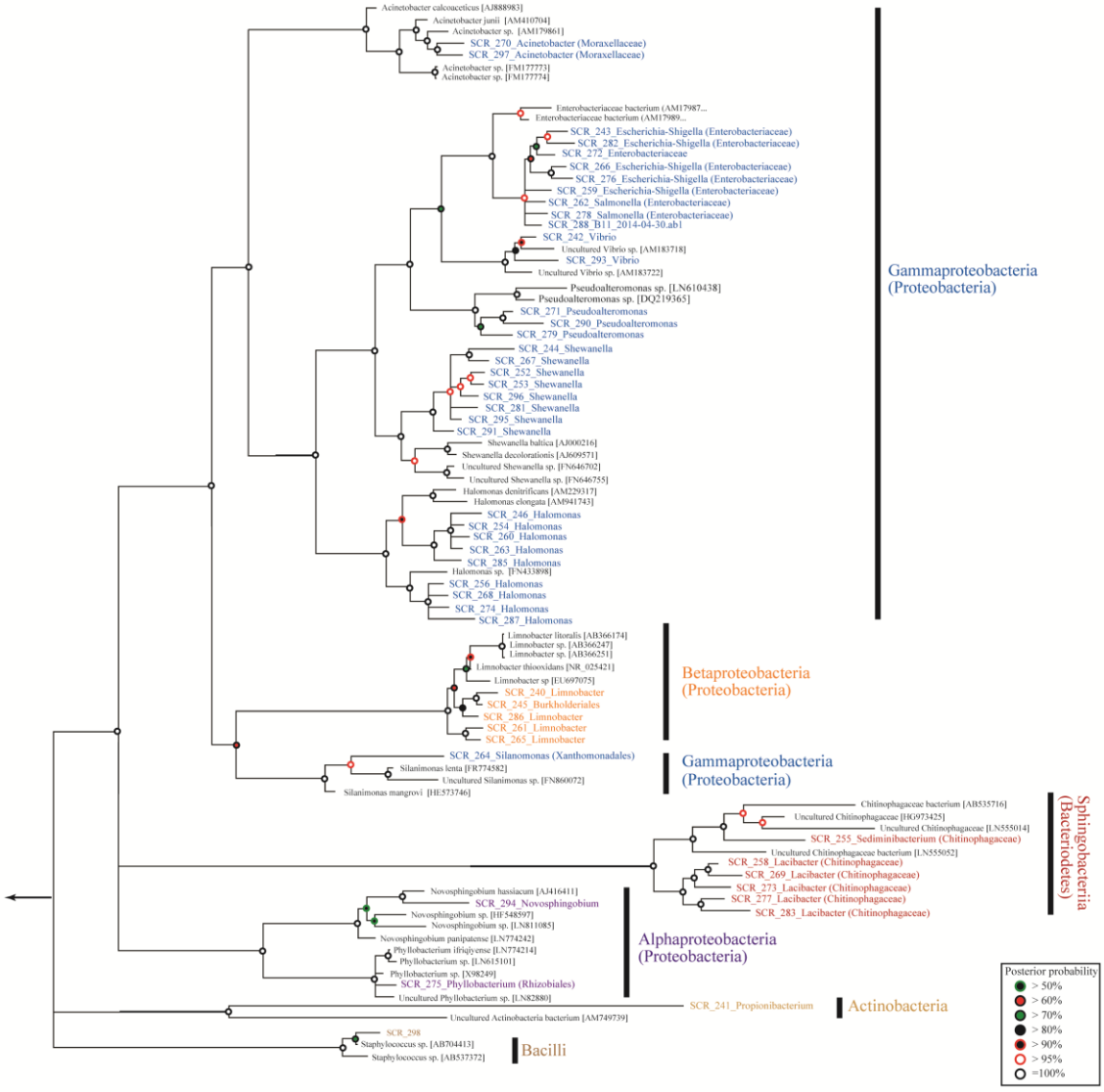


Figure C.2 Phylogenetic relationship of clones sequenced from a *P. gracilis* DNA extract clone library. The clones sequenced from the clone library are shown in colour and the NCBI reference sequences are shown in black.

A set of 27 individuals was used in a first Illumina sequencing analysis of the V3-V4 region of the 16S rRNA. The procedure for DNA extraction, amplification, sequencing and bioinformatics analysis was as described in Chapter 5. Results of the analysis showing the most abundant MED nodes (above 0.2% average relative abundance across samples) are in Figure C.3.

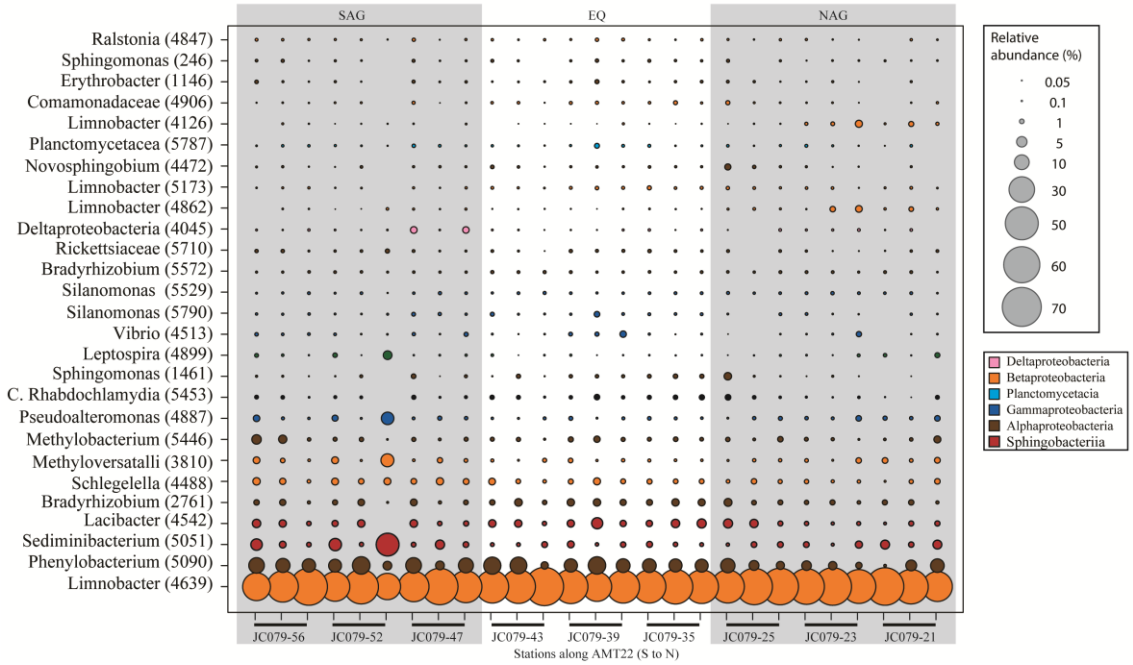


Figure C.3 Most abundant MED nodes from AMT22 Illumina sequencing dataset. The node related to Limnobacter was highly abundant in every sample.

The results of the AMT22 preliminary sequencing experiment using two different methods showed:

i. Clone library: Out of 55 successfully sequenced clones amplified from an individual female *P. gracilis* DNA extract, five (9%) of them were related to *Limnobacter* (Betaproteobacteria). The phylogenetic composition on the class level of the rest was similar to the composition of the most abundant nodes in the AMT24 dataset.

ii. Illumina sequencing: The most abundant node across samples in the AMT22 dataset was related to *Limnobacter* and the sequence 100% identical (400bp, V3-V4 region) to the most abundant *Limnobacter*-related node in the AMT24 dataset.

Appendix C.3 Specific *Limnobacter* probe design

Almost full-length 16S rRNA sequences obtained from the AMT22 clone library were used for the development of a *Limnobacter* specific FISH probe. The clone library sequences were imported into the ARB 16S rRNA reference tree and an alignment of closely related *Limnobacter* sp. sequences was used for probe design using the probe design module of ARB (Ludwig *et al.*, 2004). The probe Lim836 (5' – CAA CAA CTA GTT GAC ATC GTT TAG GG – 3') was designed, targeting position 836 of the 16S rRNA. The specificity of the probe was checked *in silico* with the online SILVA TestProbe tool (Quast *et al.*, 2013).

To determine the optimal formamide concentration during hybridisation, a formamide row test was done with a pure *Limnobacter thiooxidans* culture. The Lim836 probe was fluorescently labelled with Atto594 (Biomers, Germany). A Cy5-labelled EUB338 probe was used as a positive control and Cy3-labelled NONEUB338 probe as a negative control. The procedure was as described above.

Images showing results of probe optimisation on pure *L. thiooxidans* culture with Lim836 probe labelled with Att594 (Figure C.4) and the positive control using EUB338 labelled with Cy3 (Figure C.3).

The optimal formamide concentration of the Lim836 probe was 30% with bacterial cells clearly visible. The signal detection visibly dropped at higher formamide concentrations. Initial hybridisations with the Lim836 probe on *P. gracilis* sections were not successful and more method optimisation is needed.

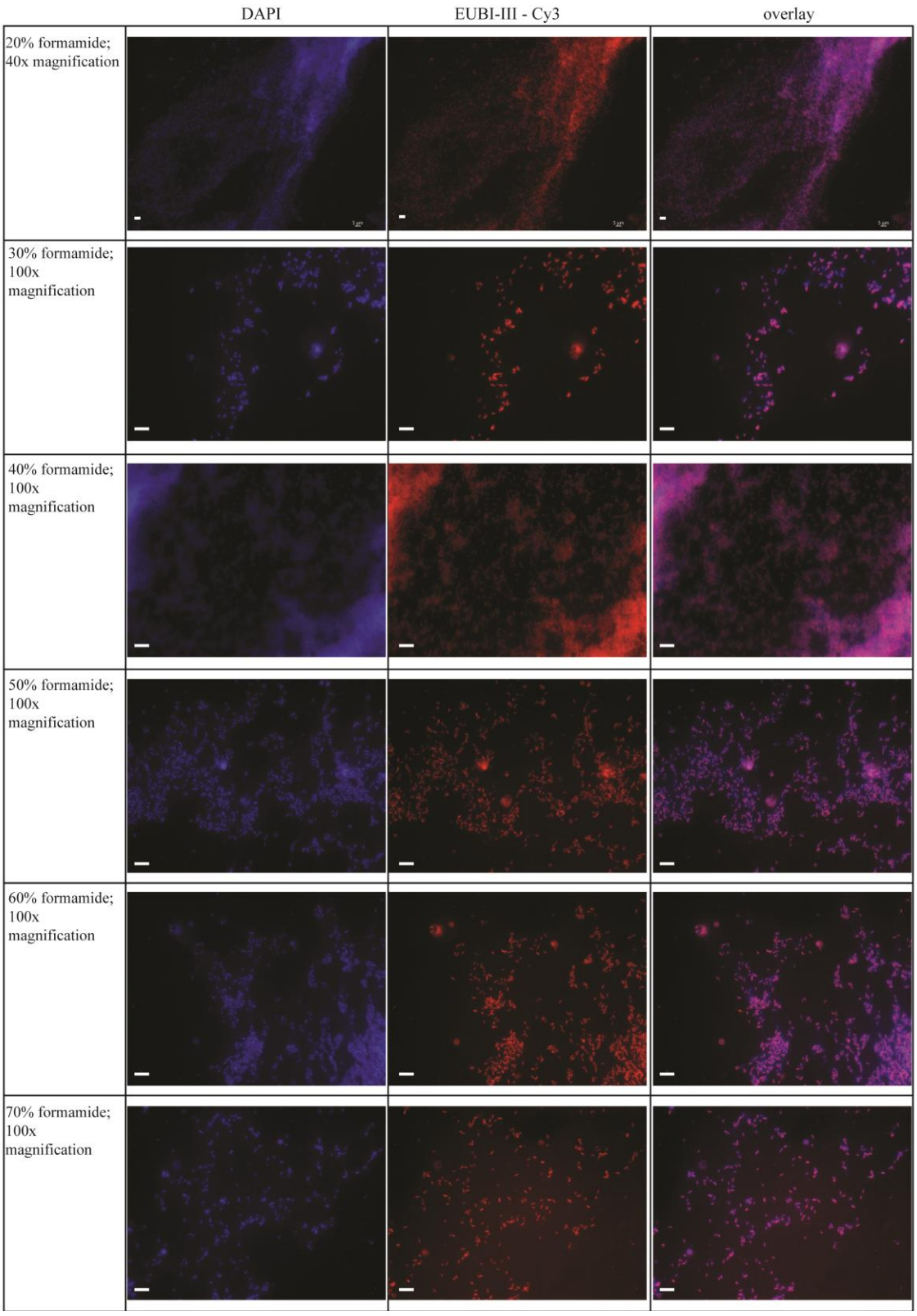


Figure C.3 Positive controls using the EUB338 probe on a pure culture of *L. thiooxidans* at varying formamide concentrations. Scale 5 um.

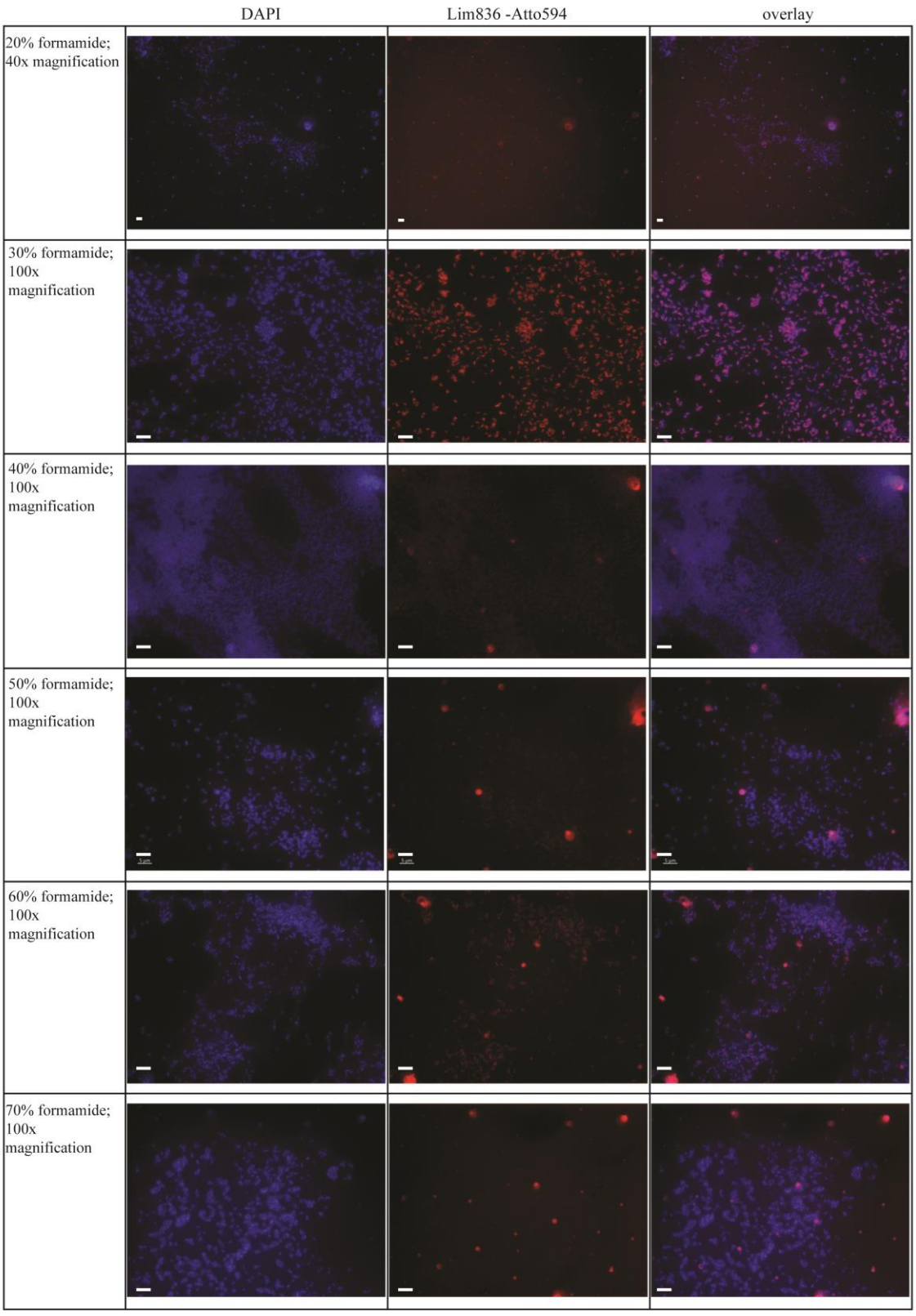


Figure C.4 Hybridisations using the Lim836 probe on a pure culture of *L. thiooxidans* at varying formamide concentrations. Scale 5 um.

List of References

Aiken, J., Rees, N, Hooker, S., Holligan, P., Bale, A., Robins, D., Moore, G., Harris, R. & Pilgrom, D. The Atlantic Meridional Transect: overview and synthesis of data. *Prog. Oceanogr.* **45,** 257–312 (2000).

Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

Amann, R. I., Krumholz, L. & Stahl, D. A. Fluorescent-olignucleotide probing of whole cells for determinitive, phylogenetic, and environmental studies in microbiology. *J. Bacteriol.* **172,** 762–770 (1990).

Amann, R. I., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59,** 143–69 (1995).

Amann, R. & Fuchs, B. M. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat. Rev. Microbiol.* **6,** 339–48 (2008).

Ambler, J. W. & Miller, C. B. Vertical habitat-partitioning by copepodites and adults of subtropical oceanic copepods. *Mar. Biol.* **94,** 561–577 (1987).

Amin, S. A., Green, D. H., Hart, M. C., Kupper, F. C., Sunda, W. G. & Carrano, C. J. Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *Proc. Natl. Acad. Sci.* **106,** 17071–17076 (2009).

Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between Diatoms and Bacteria. *Microbiol. Mol. Biol. Rev.* **76,** 667–684 (2012).

Amin, S. a. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522,** 98–101 (2015).

Angel, M. V. Biodiversity in the Pelagic Ocean. *Conserv. Biol.* **7,** 760–772 (1993).

Armougom, F. & Raoult, D. Exploring Microbial Diversity Using 16S rRNA High-Throughput Methods. *J. Comput. Sci. Syst. Biol.* **02,** 74–92 (2009).

Arnaud, J., Brunet, M. & Mazza, J. Cell and Tissue Studies on the Midgut of Cetropages typicus (Copepod, Calanoid). *Cell Tishue Res.* **187,** 333–353 (1978).

Assie, A., Borowski, C., van der Heijden, K., Raggi, L., Dubilier, N & Petersen, J. A specific and widespread epibiotic association between deep sea *Bathymodiolus* mussels and a novel family of Epsilonproteobacteria. (2016). doi:10.1111/1758-2229.12442

Avise, J. C. *et al.* Intraspecific Phylogeography : The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annu. Rev. Ecol. Syst.* **18,** 489–522 (1987).

Azam, F., Fenchel, T., Field, J.G., Graf, J. S., Meyer-Reil, L. A. & Thingstad, F. The Ecological Role of Water-Column Microbes in the Sea *. *Mar. Ecol. Prog. Ser.* **10,** 257–263 (1983).

Bannister, N. J. & Herring, P. J. Distribution and structure of luminous cells in four marine copepods. *J. Mar. Biol. Assoc. United Kingdom* **69,** 523–533 (1989).

Banse, K. Zooplankton: Pivotal role in the control of ocean production I. Biomass and production. *ICES J. Mar. Sci.* **52,** 265–277 (1995).

Baumann, P., Baumann, L., Lai, C. Y., Rouhbakhsh, D., Moran, N. A. & Clark, M. A. Genetics, physiology, and evolutionary relationships of the genus Buchnera: intracellular symbionts of aphids. *Annu. Rev. Microbiol.* **49,** 55–94 (1995).

Beaugrand, G., Ibañez, F., Lindley, J., Philip, C. & Reid, P. Diversity of calanoid copepods in the North Atlantic and adjacent seas: species associations and biogeography. *Mar. Ecol. Prog. Ser.* **232,** 179–195 (2002).

Beaugrand, G., Brander, K. M., Alistair Lindley, J., Souissi, S. & Reid, P. C. Plankton effect on cod recruitment in the North Sea. *Nature* **426,** 661–664 (2003).

Bell, M. V., Dick, J. R., Anderson, T. R. & Pond, D. W. Application of liposome and stable isotope tracer techniques to study polyunsaturated fatty acid biosynthesis in marine zooplankton. *J. Plankton Res.* **29,** 417–422 (2007).

Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing on JSTOR. *J. R. Stat. Soc. Ser. B* **57,** Vol. 57, pp. 289–300 (1995).

Berry, D., Ben Mahfoudh, K., Wagner, M. & Loy, A. Barcoded Primers Used in Multiplex Amplicon Pyrosequencing Bias Amplification. *Appl. Environ. Microbiol.* **77,** 7846–7849 (2011).

Bickel, S. L. & Tang, K. W. Carbon substrate usage by zooplankton-associated bacteria, phytoplankton-associated bacteria, and free-living bacteria under aerobic and anaerobic conditions. *Mar. Biol.* **161,** 2233–2242 (2014).

Blades-Eckelbarger, P. I. & Youngbluth, M. J. Ultrastructure of the pigment knob of Pleuromamma spp. (Copepoda: Calanoida). *J. Morphol.* **197,** 315–326 (1988).

Blanco-Bercial, L., Bradford-Grieve, J. & Bucklin, A. Molecular phylogeny of the Calanoida (Crustacea: Copepoda). *Mol. Phylogenet. Evol.* **59,** 103–13 (2011).

Blanco-Bercial, L., Cornils, A., Copley, N. & Bucklin, A. DNA Barcoding of Marine Copepods : Assessment of Analytical Approaches to Species Identification. *PLOS Curr. Tree Life* **1,** 1–21 (2014).

Boecklen, W. J., Yarnes, C. T., Cook, B. a. & James, A. C. On the Use of Stable Isotopes in Trophic Ecology. *Annu. Rev. Ecol. Evol. Syst.* **42,** 411–440 (2011).

Boon, E., Meehan, C. J., Whidden, C., Wong, D. H.-J., Langille, M. G. I. & Beiko, R. G. Interactions in the microbiome: communities of organisms and communities of genes. *FEMS Microbiol. Rev.* **38,** 90–118 (2014).

Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26,**

Braun, S. T., Proctor, L. M., Zani, S., Mellon, M. T. & Zehr, J. P. Y. Molecular evidence for zooplankton-associated nitrogen-fixing anaerobes based on amplification of the nifH gene. *FEMS Microbiol. Ecol.* **28,** 273–279 (1999).

Bright, M. & Bulgheresi, S. A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* **8,** 218–230 (2010).

Bron, J. E., Frisch, D., Goetze, E., Johnson, S. C, Lee, C. E. & Wyngaard, G. A. Observing copepods through a genomic lens. *Front. Zool.* **8,** 22 (2011).

Brucker, R. M. & Bordenstein, S. R. Speciation by symbiosis. *Trends Ecol. Evol.* **27,** 443–451 (2012).

Bucklin, A., Fros, B. W., Bradford-Grieve, J., Allen, L. D. & Copley, N. J. Molecular systematic and phylogenetic assessment of 34 calanoid copepod species of the Calanidae and Clausocalanidae. *Mar. Biol.* **142,** 333–343 (2003).

Bucklin, A. *et al.* A 'Rosetta Stone' for metazoan zooplankton: DNA barcode analysis of species diversity of the Sargasso Sea (Northwest Atlantic Ocean). *Deep Sea Res. Part II Top. Stud. Oceanogr.* **57,** 2234–2247 (2010).

Buskey, E. J., Baker, K. S., Raymond, S. C. & Elijah, S. Photosensitivity of the oceanic copepods Pleuromamma gracilis and Pleuromamma xiphias and its relationship to light penetration and daytime depth distribution *. *Mar. Ecol. Prog. Ser.* **55,** 207–216 (1989).

Calbet, A. & Saiz, E. The ciliate-copepod link in marine ecosystems. *Aquat. Microb. Ecol.* **38,** 157–167 (2005).

Calil, P. H. R., Doney, S. C., Yumimoto, K., Eguchi, K. & Takemura, T. Episodic upwelling and dust deposition as bloom triggers in low-nutrient, low-chlorophyll regions. *J. Geophys. Res. Ocean.* **116,** 1–16 (2011).

Caut, S., Angulo, E. & Courchamp, F. Variation in discrimination factors (Δ^{15}N and Δ^{13}C): the effect of diet isotopic values and applications for diet reconstruction. *J. Appl. Ecol.* **46,** 443–453 (2009).

Chandler, J. A., Lang, J., Bhatnagar, S., Eisen, J. A. & Kopp, A. Bacterial communities of diverse Drosophila species: Ecological context of a host-microbe model system. *PLoS Genet.* **7,** (2011).

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., Ellison, A. M. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84,** 45–67 (2014).

Chaston, J. & Goodrich-Blair, H. Common trends in mutualism revealed by model associations between invertebrates and bacteria. *FEMS Microbiol. Rev.* **34,** 41–58 (2010).

Chu, K. H., Li, C. P. & Ho, H. Y. The First Internal Transcribed Spacer (ITS-1) of Ribosomal DNA as a Molecular Marker for Phylogenetic and Population Analyses in Crustacea. Mar. Biotechnol. 3, 355–361 (2001).

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R. & Tiedje, J. M. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42,** 633–642 (2014).

Colman, D. R., Toolson, E. C. & Takacs-Vesbach, C. D. Do diet and taxonomy influence insect gut bacterial communities? *Mol. Ecol.* **21,** 5124–5137 (2012).

Croué, J., West, N. J., Escande, M.-L., Intertaglia, L., Lebaron, P. & Suzuki, M. T. A single betaproteobacterium dominates the microbial community of the crambescidine-containing sponge Crambe crambe. *Sci. Rep.* **3,** 2583 (2013).

Davidov, Y. & Jurkevitch, E. Diversity and evolution of Bdellovibrio-and-like organisms (BALOs), reclassification of Bacteriovorax starrii as Peredibacter starrii gen. nov., comb. nov., and description of the Bacteriovorax-Peredibacter clade as Bacteriovoracaceae fam. nov. *Int. J. Syst. Evol. Microbiol.* **54,** 1439–1452 (2004).

Davis, C. S. & McGillicuddy, D. J. Transatlantic abundance of the N2-fixing colonial cyanobacterium Trichodesmium. *Science* **312,** 1517–1520 (2006).

De Corte, D., Lekunberri, I., Sintes, E., Garcia, J. A. L., Gonzales, S. & Herndl, G. J. Linkage between copepods and bacteria in the North Atlantic Ocean. *Aquat. Microb. Ecol.* **72,** 215–225 (2014).

De Troch, M., Cnudde, C., Willems, A., Moens, T. & Vanreusel, A. Bacterial colonization on fecal pellets of harpacticoid copepods and on their diatom food. *Microb. Ecol.* **60,** 581–91 (2010).

DeNiro, M. J. & Epstein, S. Influence of diet on the distribution of nitrogen isotopes in animals. *Geochim. Cosmochim. Acta* **45,** 341–351 (1981).

DeNiro, M. J. & Epstein, S. Influence of diet on the distribution of carbon isotopes in animals. *Geochim. Cosmochim. Acta* **42,** 495–506 (1978).

Dubilier, N. The searchlight and the bucket of microbial ecology. *Environ. Microbiol.* **9,** 2–3 (2007).

Dubilier, N., Bergin, C. & Lott, C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6,** 725–40 (2008).

Dupperon, S., Bergin, C., Zielinski, F., Blazejak, A., Pernthaler, A., McKiness, Z. P., DeChaine, E., Cavanaugh, C. M. / Dubilier, N. A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia:Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Env. Microbiol.* **8**, 1441-1447 (2006).

Engel, P. & Moran, N. A. The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol. Rev.* **37,** 699–735 (2013).

Eren, A. M., Borisy, G. G., Huse, S. M. & Mark Welch, J. L. Oligotyping analysis of the human oral microbiome. *Proc. Natl. Acad. Sci.* **111,** E2875–E2884 (2014a).

Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H. & Sogin, M.L. Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* **9,** 968–979 (2014b).

Eren, A. M., Maignien, L., Sul, W. L., Murphy, L. G., Grim, S. L., Morrison, H. G. & Sogin, M. L. Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* **4,** 1111–1119 (2013).

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M. & Ravel, J. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2,** 6 (2014).

Fenchel, T. The microbial loop - 25 years later. *J. Exp. Mar. Bio. Ecol.* **366,** 99–103 (2008).

Féral, J.-P. How useful are the genetic markers in attempts to understand and manage marine biodiversity? *J. Exp. Mar. Bio. Ecol.* **268,** 121–145 (2002).

Fernandez, A., Maranon, E. & Bode, A. Large-scale meridional and zonal variability in the nitrogen isotopic composition of plankton in the Atlantic Ocean. *J. Plankton Res.* **36,** 1060–1073 (2014).

Ferrari, F. D. Pleiotropy and Pleuromamma , the Looking-Glass Copepods (Calanoida). *Crustac. Stud. Copepoda II (Proceedings First Int. Conf. Copepoda, Amsterdam, Netherlands, 24-28 August 1981* 166–181 (1984).

Ferrari, F. & Fornshell, J. Morphological variability of Pleuromamma abdominalis (Copepoda, Calanoida, Metridinidae) along two latitudinal transects in the eastern North Pacific Ocean. *Crustaceana* **83,** 753–765 (2010).

Field, C. B., Behrenfel, M. J., Randerson, J. T. & Falkowski, P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281,** 237–240 (1998).

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3,** 294–9 (1994).

France, R. L. Carbon-13 enrichment in benthic compared to planktonic algae: foodweb implications. *Mar. Ecol. Prog. Ser.* **124,** 307–312 (1995).

Frangoulis, C., Christou, E. D. & Hecq, J. H. Comparison of marine copepod outfluxes: nature, rate, fate and role in the carbon and nitrogen cycles. *Adv. Mar. Biol.* **47,** (2005).

Frank, D. N. BARCRAWL and BARTAB: software tools for the design and implementation of barcoded primers for highly multiplexed DNA sequencing. *BMC Bioinformatics* **10,** 362 (2009).

Franzenburg, S., Fraune, S., Altrock, P. M., Künzel, S., Baines, J. F., Traulsen, A. & Bosch, T. C. G. Bacterial colonization of Hydra hatchlings follows a robust temporal pattern. *ISME J.* **7,** 781–90 (2013).

Freese, H. M. & Schink, B. Composition and stability of the microbial community inside the digestive tract of the aquatic crustacean Daphnia magna. *Microb. Ecol.* **62,** 882–94 (2011).

Gallienne, C. P. & Robins, D. B. Is Oithona the most important copepod in the world's oceans ? *J. Plankton Res.* **23,** 1421–1432 (2001).

Gerdts, G., Brandt, P., Kreisel, K., Boersma, M., Schoo, K. L. & Wichels, A. The microbiome of North Sea copepods. *Helgol Mar Res* **67,** 757–773 (2013).

Givens, C. E., Burnett, K. G., Burnett, L. E. & Hollibaugh, J. T. Microbial communities of the carapace, gut , and hemolymph of the Atlantic blue crab , Callinectes sapidus. *Mar. Biol.* **160,** 2841–2851 (2013)

Goericke, R. & Fry, B. Variations of marine plankton δ 13 C with latitude, temperature, and dissolved CO 2 in the world ocean. *Global Biogeochem. Cycles* **8,** 85–90 (1994).

Goetze, E. Population Differentiation in the Open Sea: Insights from the Pelagic Copepod Pleuromamma xiphias. *Integr. Comp. Biol.* **51,** 580–597 (2011).

Gómez-Pereira, P. R., Hartmann, M., Grob, C., Tarran, G. A., Martin, A. P., Fuchs, B. M., Scanlan, D. J. & Zubkov, M. V. Comparable light stimulation of organic nutrient uptake by SAR11 and Prochlorococcus in the North Atlantic subtropical gyre. *ISME J.* **7,** 603–14 (2013).

Goswami, S. C., Gajbhiye, S. N. & Padmavati, G. Distribution of Pleuromamma (Copepoda: Metridiidae) along a north-south transect in the Indian Ocean. *Oceanogr. Indian Ocean* 157–166 (1992).

Graham, B. S., Koch, P. L., Newsome, S. D., Mcmahon, K. W. & Aurioles, D. *Isoscapes*. (Springer Netherlands, 2010). doi:10.1007/978-90-481-3354-3

Grossart, H.-P., Dziallas, C. & Tang, K. W. Bacterial diversity associated with freshwater zooplankton. *Environ. Microbiol. Rep.* **1,** 50–55 (2009).

Großkopf, T. *et al.* Doubling of marine dinitrogen-fixation rates based on direct measurements. *Nature* **488,** 361–364 (2012).

Gruber-Vodicka, H. R., Dirks, U., Leisch, N., Baranyi, C., Stoecker, K., Bulgheresi, S., Heindl, N. R., Horn, M., Lott, C., Loy, A., Wagner, M. & Ott, J. Paracatenula, an ancient symbiosis between thiotrophic Alphaproteobacteria and catenulid flatworms. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 12078–83 (2011).

Haddock, S. H. D., Moline, M. A. & Case, J. F. Bioluminescence in the sea. *Ann. Rev. Mar. Sci.* **2,** 443–93 (2010).

Halbert, K. M. K., Goetze, E. & Carlon, D. B. High Cryptic Diversity across the Global Range of the Migratory Planktonic Copepods Pleuromamma piseki and P . gracilis. *PLoS One* **8,** (2013).

Hallberg, E. & Hirche, H.-J. Diferentiation of mid-gut in adults and over-wintering copepodids of Calanus finmarchicus (Gunnerus) and C. helgolandicus Claus. *J. Exp. Mar. Bio. Ecol.* **48,** 283–295 (1980).

Hansen, B. & Bech, G. Bacteria associated with a marine planktonic copepod in culture . I . Bacterial genera in seawater , body surface , intestines and fecal pellets and succession during fecal pellet degradation. *J. Plankton Res.* **18,** 257–273 (1996).

Har, J. Y., Helbig, T., Lim, J. H., Fernando, S. C., Reitzel, A. M., Penn, K. & Thompson, J. R. Microbial diversity and activity in the Nematostella vectensis holobiont: insights from 16S rRNA gene sequencing, isolate genomes, and a pilot-scale survey of gene expression. *Front. Microbiol.* **6,** (2015).

Harper, G. L., King, R. A., Dodd, C. S., Harwood, J. D., Glen, D. M. & Bruford, M. W. Rapid screening of invertebrate predators for multiple prey DNA targets. *Mol. Ecol.* **14,** 819–827 (2005).

Harris, J. M. The Presence, Nature, and Role of Gut Microflora in Aquatic Invertebrates: A Synthesis. *Microb. Ecol.* **25,** 195–231 (1993).

Hartmann, M., Grob, C., Tarran, G. A., Martin, A. P., Burkill, P. H., Scanlan, D. J. & Zubkov, M. V. Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proc. Natl. Acad. Sci.* **109,** 5756–5760 (2012).

Hassett, R. P. & Blades-Eckelbarger, P. Diel changes in gut-cell morphology and digestive activity of the marine copepod Acartia tonsa. *Mar. Biol.* **124,** 59–69 (1995).

Heidelberg, J. F., Heidelberg, K. B. & Colwell, R. R. Bacteria of the γ -Subclass Proteobacteria Associated with Zooplankton in Chesapeake Bay. *Appl. Environ. Microbiol.* **68,** 5498–5507 (2002).

Heywood, J. L., Zubkov, M. V., Tarran, G. A., Fuchs, B. M. & Holligan, P. M. Prokaryoplankton standing stocks in oligotrophic gyre and equatorial provinces of the Atlantic Ocean: Evaluation of inter-annual variability. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **53,** 1530–1547 (2006).

Hirai, J., Tsuda, A. & Goetze, E. Extensive genetic diversity and endemism across the global range of the oceanic copepod Pleuromamma abdominalis. *Prog. Oceanogr.* **138,** 77–90 (2015).

Hood, R. R., Coles, V. J. & Capone, D. G. Modeling the distribution of Trichodesmium and nitrogen fixation in the Atlantic Ocean. *J. Geophys. Res.* **109,** (2004).

Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17,** 754–755 (2001).

Huq, A., Small, E. B., West, P. A., Huq, M. I., Rahman, R. & Colwell, R. R. Ecological relationships between Vibrio cholerae and planktonic crustacean. *Appl. Environ. Microbiol.* **45,** 275–283 (1983).

Ide, K., Takahashi, K., Kuwata, a., Nakamachi, M. & Saito, H. A rapid analysis of copepod feeding using FlowCAM. *J. Plankton Res.* **30,** 275–281 (2007).

Ikeda, T., Yamaguchi, A. & Matsuishi, T. Chemical composition and energy content of deep-sea calanoid copepods in the Western North Pacific Ocean. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **53,** 1791–1809 (2006).

Irwin, A. J. & Oliver, M. J. Are ocean deserts getting larger? *Geophys. Res. Lett.* **36,** 1–5 (2009).

Iversen, M. H. & Poulsen, L. K. Coprorhexy, coprophagy, and coprochaly in the copepods Calanus helgolandicus, Pseudocalanus elongatus, and Oithona similis. *Mar. Ecol. Prog. Ser.* **350,** 79–89 (2007).

Jackson, A. L., Inger, R., Parnell, A. C. & Bearhop, S. Comparing isotopic niche widths among and within communities: SIBER - Stable Isotope Bayesian Ellipses in R. *J. Anim. Ecol.* **80,** 595–602 (2011).

Jardillier, L., Zubkov, M. V, Pearman, J. & Scanlan, D. J. Significant CO2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J.* **4,** 1180–1192 (2010).

Kaltenpoth, M., Yildirim, E., Gürbüz, M. F., Herzner, G. & Strohm, E. Refining the roots of the beewolf-streptomyces symbiosis: Antennal symbionts in the rare genus Philanthinus (Hymenoptera, Crabronidae). *Appl. Environ. Microbiol.* **78,** 822–827 (2012).

Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).

Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).

Kiers, T. & West, S. Evolving new organisms via symbiosis. *Science* **348,** 392–393 (2015).

Kikuchi, Y., Meng, X. & Fukatsu, T. Gut Symbiotic Bacteria of the Genus Burkholderia in the Broad-Headed Bugs Riptortus clavatus and Leptocorisa chinensis (Heteroptera : Alydidae). *Appl. Environ. Microbiol.* **71,** 4035–4043 (2005).

Kikuchi, Y., Hosokawa, T. & Fukatsu, T. Insect-microbe mutualism without vertical transmission: A stinkbug acquires a beneficial gut symbiont from the environment every generation. *Appl. Environ. Microbiol.* **73,** 4308–4316 (2007).

King, R. A, Read, D. S., Traugott, M. & Symondson, W. O. C. Molecular analysis of predation: a review of best practice for DNA-based approaches. *Mol. Ecol.* **17,** 947–63 (2008).

Kleppel, G. S. On the diets of calanoid copepods. *Mar. Ecol. Prog. Ser.* **99,** 183–195 (1993).

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. & Glöckner, F. O. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41,** 1–11 (2013).

Laakmann, S. & Auel, H. Longitudinal and vertical trends in stable isotope signatures (13 C and 15 N) of omnivorous and carnivorous copepods across the South Atlantic Ocean. *Mar. Biol.* **157,** 463–471 (2010).

Lampitt, R. S., Noji, T. & von Bodungen, B. What happens to zooplankton faecal pellets? Implications for material flux. *Mar. Biol.* **104,** 15–23 (1990).

Landrum, J. P., Altabet, M. a. & Montoya, J. P. Basin-scale distributions of stable nitrogen isotopes in the subtropical North Atlantic Ocean: Contribution of diazotroph nitrogen to particulate organic matter and mesozooplankton. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **58,** 615–625 (2011).

Landry, M. R., Al-Mutairi, H., Selph, K. E., Christensen, S. & Nunnery, S. Seasonal patterns of mesozooplankton abundance and biomass at Station ALOHA. *Deep. Res. Part II Top. Stud. Oceanogr.* **48,** 2037–2061 (2001).

Layman, C. A., Arrington, D. A., Montaña, C. G. & Post, D. M. Can stable isotope ratios provide for community-wide measures of trophic structure? *Ecology* **88,** 42–48 (2007).

Lemaitre, B. & Hoffmann, J. The host defense of Drosophila melanogaster. *Annu. Rev. Immunol.* **25,** 697–743 (2007).

Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320,** 1647–1651 (2008).

Lidstrom, M. E. in *The Prokaryotes* 618–634 (Springer New York, 2006

Loman, N. J.*, Misra, R. V., Dallman, T. J.,* Constantinidou, C., Gharbia, S. E., Wain, J. & Pallen, M. J. Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30,** 434–9 (2012).

Longhurst, A. R. & Glen Harrison, W. The biological pump: Profiles of plankton production and consumption in the upper ocean. *Prog. Oceanogr.* **22,** 47–123 (1989).

López, E., Anadón, R. & Harris, R. P. Functional responses of copepod nauplii using a high efficiency gut fluorescence technique. *Mar. Biol.* **150,** 893–903 (2007).

López, E. & Anadón, R. Copepod communities along an Atlantic Meridional Transect: Abundance, size structure, and grazing rates. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **55,** 1375–1391 (2008).

Loy, A., Arnold, R., Tischler, P., Rattei, T., Wagner, M. & Horn, M. ProbeCheck - A central resource for evaluating oligonucleotide probe coverage and specificity. *Environ. Microbiol.* **10,** 2894–2898 (2008).

Lu, H., Sato, Y., Fujimura, R., Nishizawa, T., Kamijo, T. & Ohta, H. Limnobacter litoralis sp. nov., a thiosulfate-oxidizing, heterotrophic bacterium isolated from a volcanic deposit, and

emended description of the genus Limnobacter. *Int. J. Syst. Evol. Microbiol.* **61,** 404–407 (2011).

Machida, R. J., Miya, M. U., Nishida, M. & Nishida, S. Large-scale gene rearrangements in the mitochondrial genomes of two calanoid copepods Eucalanus bungii and Neocalanus cristatus (Crustacea), with notes on new versatile primers for the srRNA and COI genes. *Gene* **332,** 71–8 (2004).

Machida, R. J., Miya, M. U., Nishida, M. & Nishida, S. Molecular phylogeny and evolution of the pelagic copepod genus Neocalanus (Crustacea: Copepoda). *Mar. Biol.* **148,** 1071–1079 (2005).

Manz, W., Amann, R., Ludwig, W., Wagner, M. & Schleifer, K.-H. Phylogenetic Oligodeoxynucleotide Probes for the Major Subclasses of Proteobacteria: Problems and Solutions. *Syst. Appl. Microbiol.* **15,** 593–600 (1992).

Marañón, E., Holligan, P. M., Barciela, R., Gonzales, N., Mouriño, B., Pazó, M. J. & Varela, M. Patterns of phytoplankton size structure and productivity in contrasting open-ocean environments. *Mar. Ecol. Prog. Ser.* **216,** 43–56 (2000).

Marañón, E. *et al.* Patterns of phytoplankton size structure and productivity in contrasting open-ocean environments. *Mar. Ecol. Prog. Ser.* **216,** 43–56 (2001).

Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–80 (2005).

Martinson, V. G., Moy, J. & Moran, N. a. Establishment of characteristic gut bacteria during development of the honeybee worker. *Appl. Environ. Microbiol.* **78,** 2830–40 (2012).

McCutchan, J. H., Lewis, W. M., Kendall, C. & McGrath, C. C. Variation in Trophic Shift for Stable Isotope Ratios of Carbon , Nitrogen , and Sulfur Variation in trophic shift for stable isotope ratios of carbon , nitrogen , and sulfur. *OIKOS* **102,** 378–390 (2003).

McDonald, K. L. Rapid Embedding Methods into Epoxy and LR White Resins for Morphological and Immunological Analysis of Cryofixed Biological Specimens. *Microsc. Microanal.* **20,** 152–163 (2014).

McFall-Ngai, M. J. Consequences of evolvingwith bacterial symbionts: Insights from the Squid-Vibrio Associations. *Annu. Rev. Ecol. Syst.* **30,** 235–56 (1999).

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Nealson, K., Pierce, N. E., Rawls, J. F., Reid, A., Ruby, E. G., Rumpho, M., Sanders, J. G., Tautz, D. & Wernegreen, J. J. Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci.* **110,** 3229–3236 (2013).

Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11,** 31–46 (2010).

Mereschkowsky K.S. (1910). "Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Ent-stehung der Organismen". Biol Centralbl **30**: 353-367

Michels, J. & Gorb, S. N. Mandibular gnathobases of marine planktonic copepods -feeding tools with complex micro- and nanoscale composite architectures. *Beilstein J. Nanotechnol.* **6,** 674–685 (2015).

Michener, R. & Lajtha, K. *Stable isotopes in ecology and environmental science*. (Blackwell Publishing, 2007).

Minagawa, M. & Wada, E. Stepwise enrichment of 15N along food chains: Further evidence and the relation between δ15N and animal age. *Geochim. Cosmochim. Acta* **48,** 1135–1140 (1984).

Mino, Y., Saino, T., Suzuki, K. & Marañón, E. Isotopic composition of suspended particulate nitrogen (δ 15 N sus) in surface waters of the Atlantic Ocean from 50°N to 50°S. *Global Biogeochem. Cycles* **16,** 7–1–7–9 (2002).

Møller, E. F. & Nielsen, T. G. Production of Bacterial Substrate by Marine Copepods: Effect of Phytoplankton Biomass and Cell Size. *J. Plankton Res.* **23,** 527–536 (2001).

Møller, E. F., Riemann, L. & Søndergaard, M. Bacteria associated with copepods : abundance , activity and community composition. *Aquat. Microb. Ecol.* **47,** 99–106 (2007).

Montoya, J. P., Carpenter, E. J. & Capone, D. G. Nitrogen fixation and nitrogen isotope abundances in zooplankton of the oligotrophic North Atlantic. *Limnol. Oceanogr.* **47,** 1617–1628 (2002).

Montoya, J.P. Natural abundance of 15N in marine planktonic ecosystems. *Stable isotopes in ecology and environmental science*. p. 176. (2007).

Moore, M. C. *et al.* Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nat. Geosci.* **2,** 867–871 (2009).

Moore, C. M. *et al.* Processes and patterns of oceanic nutrient limitation. *Nat. Geosci* **6,** 701–710 (2013).

Morales, C. E. Carbon and nitrogen fluxes in the oceans: the contribution by zooplankton migrants to active transport in the North Atlantic during the Joint Global Ocean Flux Study. *J. Plankton Res.* **21,** 1799–1808 (1999).

Morris, M. & Hopkins, T. Biochemical composition of crustacean zooplankton from the eastern Gulf of Mexico. *J. Exp. Mar. Bio. Ecol.* **69,** 1–19 (1983).

Moya, A., Peretó, J., Gil, R. & Latorre, A. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* **9,** 218–229 (2008).

Nybakken JW, Bertness MD (2004) Marine Biology: an ecological approach. 6th edition, pp. 144-195. Pearson Education, Inc., publishing as Benjamin Cummings, San Francisco.

Nagasawa, S. & Nemoto, T. Presence of bacteria in guts of marine crustaceans and on their fecal pellets. *J. Plankton Res.* **10,** 559–564 (1988).

Nejstgaard, J. C., Frischer, J. C., Raule, C. L., Gruebel, R., Kohlberg, K. E. & Verity, P. G. Molecular detection of algal prey in copepod guts and fecal pellets. *Limnol. Oceanogr. Methods* **1,** 29–38 (2003).

Nejstgaard, J. C., Frischer, M. E., Simonelli, P., Troedsson, C., Brakel, M., Adiyaman, F., Sazhin, A. F. & Artigas, L. F. Quantitative PCR to estimate copepod feeding. *Mar. Biol.* **153,** 565–577 (2007).

Newton, J. in *Encycl. Life Sci.* (John Wiley & Sons, Ltd, 2010). doi:10.1002/9780470015902.a0021231

Niiler. (2001). The global flow field. In "ocean circulation and climate" (G. Siedler, J. Church, and J. Gould, Eds.). Elsevier

Oksanen, J. *et al.* Package ' vegan '. (2016).

Ott, J. A. *et al.* Tackling the Sulfide Gradient : A Novel Strategy Involving Marine Nematodes and Chemoautotrophic Ectosymbionts. *Mar. Ecol.* **12,** 261–279 (1991).

Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276,** 734–740 (1997).

Parnell, A. C., Inger, R., Bearhop, S. & Jackson, A. L. Source Partitioning Using Stable Isotopes: Coping with Too Much Variation. *PLoS One* **5,** e9672 (2010).

Pedrós-Alió, C. Marine microbial diversity: can it be determined? *Trends Microbiol.* **14,** 257–263 (2006).

Peerakietkhajorn, S., Tsukada, K., Kato, Y., Matsuura, T. & Watanabe, H. Symbiotic bacteria increase the population size of a freshwater crustacean , Daphnia magna. **7,** 2011 (2011).

Perez-Cobas, A. E., Maiques, E., Angelova, A., Carrasco, P., Moya, A. & Latorre, A. Diet shapes the gut microbiota of the omnivorous cockroach Blatella germanica. *FEMS Microbiol. Ecol.* 1–14 (2015).

Petersen, J. M., Ramette, A., Lott, C., Cambon-Bonavita, M-A., Zbinden, M. & Dubilier, N. Dual symbiosis of the vent shrimp Rimicaris exoculata with filamentous gamma- and epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environ. Microbiol.* **12,** 2204–18 (2010).

Polovina, J. J., Howell, E. A. & Abecassis, M. Ocean's least productive waters are expanding. *Geophys. Res. Lett.* **35,** 2–6 (2008).

Polz, M. F., Felbeck, H., Novak, R., Nebelsick, M. & Ott, J. A. Chemoautotrophic, sulfur-oxidizung symbiontic bacteria on marine nematodes: morphological and biochemical characterization. *Microb. Ecol.* **24,** 313–329 (1992).

Pond, D. W. & Ward, P. Importance of diatoms for Oithona in Antarctic waters. *J. Plankton Res.* **33,** 105–118 (2011).

Popp, B. N., Laws, E. A., Bidigare, R. R., Dore, J. E., Hanson, K. L. & Wakeham, S. G. Effect of Phytoplankton Cell Geometry on Carbon Isotopic Fractionation. *Geochim. Cosmochim. Acta* **62,** 69–77 (1998).

Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14,** 817–818 (1998).

Post, D. M. Using stable isotopes to estimate trophic position: models, methos, and assumptions. . *Ecology* **83,** 703–718 (2002).

Proctor, L. M. Nitrogen-fixing , photosynthetic , anaerobic bacteria associated with pelagic copepods. *Aquat. Microb. Ecol.* **12,** 105–113 (1997).

Pruesse, E., Peplies, J. & Glockner, F. O. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28,** 1823–1829 (2012).

Quast, C. Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41,** D590–D596 (2013).

Ramette, A. Multivariate analyses in microbial ecology. *FEMS Microbiol. Ecol.* **62,** 142–160 (2007).

Rau, G. H., Sweeney, R. E. & Kaplan, I. R. Plankton 13C:12C ratio changes with latitude: differences between northern and southern oceans. *Deep Sea Res. Part A. Oceanogr. Res. Pap.* **29,** 1035–1039 (1982).

Rau, G. H., Riebesell, U. & Wolf-Gladrow, D. A model of photosynthetic 13C fractionation by marine phytoplankton based on diffusive molecular CO2 uptake. *Mar. Ecol. Prog. Ser.* **133,** 275–285 (1996).

Rees, A., Robinson, C., Smyth, T., Aiken, J., Nightingale, P. & Zubkov, M. 20 Years of the Atlantic Meridional Transect-AMT. *Limnol. Oceanogr. Bull.* **24,** 101–107 (2015).

Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics. Proteomics Bioinformatics* **13,** 278–289 (2015).

Rossello-Mora, R. & Amann, R. The species concept for prokayotes . *FEMS Microbiol. Rev.* **25 ,** 39–67 (2001).

Ruby, E. G. & McFall-Ngai, M. J. A squid that glows in the night: Development of an animal-bacterial mutualism. *J. Bacteriol.* **174,** 4865–4870 (1992).

Sanders, J. G., Beinart, R. a, Stewart, F. J., Delong, E. F. & Girguis, P. R. Metatranscriptomics reveal differences in in situ energy and nitrogen metabolism among hydrothermal vent snail symbionts. *ISME J.* **7,** 1556–67 (2013).

Scavotto, R. E., Dziallas, C., Bentzon-Tilia, M., Riemann, L. & Moisander, P. H. Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean. *Environ. Microbiol.* **17**, 3754-3765 (2015).

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T. & Quince, C. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43,** 1–16 (2015).

Schirmer, M., D'Amore, R., Ijaz, U. Z., Hall, N. & Quince, C. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* **17,** 125 (2016).

Schizas, N. V., Street, G. T., Coull, B. C., Chandler, G. T., Quattro, J. M. Molecular population structure of the marine benthic copepod *Microarthridion littorale* along the southeastern and Gulf coasts of nthe USA. *Marine Biology* **135**, 399 (1999).

Schloss, P. D. *et al.* Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).

Schloss, P. D., Gevers, D. & Westcott, S. L. Reducing the effects of PCR amplification and sequencing Artifacts on 16s rRNA-based studies. *PLoS One* **6,** (2011).

Schnack-Schiel, S. B., Mizdalski, E. & Cornils, A. Copepod abundance and species composition in the Eastern subtropical/tropical Atlantic. *Deep. Res. Part II Top. Stud. Oceanogr.* **57,** 2064–2075 (2010).

Seguritan, V. & Rohwer, F. FastGroup: a program to dereplicate libraries of 16S rDNA sequences. *BMC Bioinformatics* **2,** 9 (2001).

Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–1145 (2008).

Shoemaker, K. M. & Moisander, P. H. Microbial diversity associated with copepods in the North Atlantic subtropical gyre. *FEMS Microbiol. Ecol.* 1–11 (2015).

Smyth, T. *AMT24 Cruise Report*. (2014).

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M. & Herndl, G. J. Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc. Natl. Acad. Sci.* **103,** 12115–12120 (2006).

Sohm, J. A., Hilton, J. A., Noble, A. E., Zehr, J. P., Saito, M. A. & Webb, E. A. Nitrogen fixation in the South Atlantic Gyre and the Benguela Upwelling System. *Geophys. Res. Lett.* **38,** (2011).

Song, H., Buhay, J. E., Whiting, M. F. & Crandall, K. A. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci.* **105,** 13486–13491 (2008).

Sommer, U., Stibor, H., Katechakis, a, Sommer, F. & Hansen, T. Pelagic food web congurations at different levels of nutrient richness and their implications for the ratio sh production: primary production. *Hydrobiologia* **484,** 11–20 (2002).

Spring, S., Kampfer, P. & Schleifer, K. H. Limnobacter thiooxidans gen. nov., sp. nov., a novel thiosulfate- oxidizing bacterium isolated from freshwater lake sediment. *Int J Syst Evol Microbiol* **51,** 1463–70. (2001).

Steinberg, D. K., Carlson, C. A., Bates, N. R., Goldthwait, S. A., Madin, L. P., Madin, L. P. & Michaels, A. F. Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **47,** 137–158 (2000).

Steur A (1932) Copepoda (6): Pleuromamma GIesbr. 1898 der Deutschen Tiefsee-Expedition. Wissenschaftliche Ergebnisse der Deutschen Tiefsee-Expe- dition auf em Dampfer ‘‘Valdivia’’ 1898–1899 24: 1–119.

Stoecker, D. K. & Capuzzo, J. M. Predation on Protozoa: Its importance to zooplankton. *J. Plankton Res.* **12,** 891–908 (1990).

Tang, K. W. Copepods as microbial hotspots in the ocean: effects of host feeding activities on attached bacteria. *Aquat. Microb. Ecol.* **38,** 31–40 (2005).

Tang, K., Dziallas, C., Hutalle-Schmelzer, K. & Grossart, H.-P. Effects of food on bacterial community composition associated with the copepod Acartia tonsa Dana. *Biol. Lett.* **5,** 549–53 (2009).

Tang, K. W., Turk, V. & Grossart, H.-P. Linkage between crustacean zooplankton and aquatic bacteria. *Aquat. Microb. Ecol.* **61,** 261–277 (2010).

Tang, K. W., Dziallas, C. & Grossart, H.-P. Zooplankton and aggregates as refuge for aquatic bacteria: protection from UV, heat and ozone stresses used for water treatment. *Environ. Microbiol.* **13,** 378–90 (2011).

Tarran, G. A., Heywood, J. L. & Zubkov, M. V. Latitudinal changes in the standing stocks of nano- and picoeukaryotic phytoplankton in the Atlantic Ocean. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **53,** 1516–1529 (2006).

Teuber, L., Schukat, A., Hagen, W. & Auel, H. Trophic interactions and life strategies of epi- to bathypelagic calanoid copepods in the tropical Atlantic Ocean. *J. Plankton Res.* **36,** 1109–1123 (2014).

Tholen, A., Schink, B. & Brune, A. The gut microflora of Reticulitermes flavipes, its relation to oxygen, and evidence for oxygen-dependent acetogenesis by the most abundant Enterococcus sp. *FEMS Microbiol. Ecol.* **24,** 137–149 (2006).

Tomczak, M., and Godfrey, S. J. (2003). The Atlantic Ocean. In "Regional Oceanography: an introduction", pp. 229-252. Daya publishing house, Delhi.

Tsuchida, T., Koga, R. & Fukatsu, T. Host plant specialization governed by facultative symbiont. *Science* **303,** 1989 (2004).

Turner, J. T. The Importance of Small Pelagic Planktonic Copepods and Their Role in Pelagic Marine Food Webs. *Zool. Stud.* **43,** 255–266 (2004).

Tyrrell, T., Marañón, E., Poulton, A. J., Bowie, A. R., Harbour, D. S. & Woodward, E. M. S. Large-scale latitudinal distribution of Trichodesmium spp . in the Atlantic Ocean. *J. Plankton Res.* **25,** 405–416 (2003).

van Elsas, J. D. & Boersma, F. G. H. A review of molecular methods to study the microbiota of soil and the mycosphere. *Eur. J. Soil Biol.* **47,** 77–87 (2011).

Vander Zanden, M. J., Cabana, G. & Rasmussen, J. B. Comparing trophic position of freshwater fish calculated using stable nitrogen isotope ratios (δ15N) and literature dietary data. *Can. J. Fish. Aquat. Sci.* **54,** 1142–1158 (1997).

Vanderklift, M. A. & Ponsard, S. Sources of variation in consumer-diet ?15N enrichment: a meta-analysis. *Oecologia* **136,** 169–182 (2003).

Verity, P. G. & Smetacek, V. Organism life cycles , predation , and the structure of marine pelagic ecosystems. *Mar. Ecol. Prog. Ser.* **130,** 277–293 (1996).

Verity, P. G., Smetacek, V. & Smayda, T. J. Status, trends and the future of the marine pelagic ecosystem. *Environ. Conserv.* **29,** 207–237 (2002).

Wallner, G., Amann, R. & Beisker, W. Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* **14,** 136–143 (1993).

Williams, H. N. *et al.* Halobacteriovorax, an underestimated predator on bacteria: potential impact relative to viruses on bacterial mortality. *ISME J.* **10,** 1–9 (2015).

Wilson, S. E. & Steinberg, D. K. Autotrophic picoplankton in mesozooplankton guts: Evidence of aggregate feeding in the mesopelagic zone and export of small phytoplankton. *Mar. Ecol. Prog. Ser.* **412,** 11–27 (2010).

Woese, C. R. Bacterial Evolution. *Microbiology* **51,** 221–271 (1987).

Zehr, J. P. & Kudela, R. M. Nitrogen Cycle of the Open Ocean: From Genes to Ecosystems. *Ann. Rev. Mar. Sci.* **3,** 197–225 (2011).

Zhou, J., Bruns, M. A. N. N. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62,** 316–322 (1996).

Zhou, J. *et al.* Reproducibility and quantitation of amplicon sequencing-based detection. *ISME J.* **5,** 1303–1313 (2011).

Zilber-Rosenberg, I. & Rosenberg, E. Role of microorganisms in the evolution of animals and plants: The hologenome theory of evolution. *FEMS Microbiol. Rev.* **32,** 723–735 (2008).

Zimmermann, J. *et al.* Dual symbiosis with co-occurring sulfur-oxidizing symbionts in vestimentiferan tubeworms from a Mediterranean hydrothermal vent. *Environ. Microbiol.* (2014). doi:10.1111/1462-2920.12427

Zimmermann, J., Wentrup, C., Sadowski, M., Blazejak, A., Gruber-Vodicka, H., Kleiner, M., Ott, J., Cronholm, B., De Wit, P., Erseus, C. & Dubilier, N. Closely coupled evolutionary

history of ecto- and endosymbionts from two distantly-related animal phyla. *Mol. Ecol.* (2016).

Zubkov, M. V., Sleigh, M. A., Tarran, G. A., Burkill, P. H. & Leakey, R. J. G. Picoplanktonic community structure on an Atlantic transect from 50N to 50S. *Deep. Res. Part I Oceanogr. Res. Pap.* **45,** 1339–1355 (1998).

Zubkov, M. V., Sleigh, M. A., Burkill, P. H. & Leakey, R. J. G. Picoplankton community structure on the Atlantic Meridional Transect: a comparison between seasons. *Prog. Oceanogr.* **45,** 369–386 (2000).

Zubkov, M. V. & López-Urrutia, A. Effect of appendicularians and copepods on bacterioplankton composition and growth in the English Channel. *Aquat. Microb. Ecol.* **32,** 39–46 (2003).

Zubkov, M. V & Tarran, G. a. High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature* **455,** 224–6 (2008).

Zubkov, M. V., Martin, A. P., Hartmann, M., Grob, C. & Scanlan, D. J. Dominant oceanic bacteria secure phosphate using a large extracellular buffer. *Nat. Commun.* **6,** 7878 (2015).

Zwirglmaier, K., Reid, W. D. K., Heywood, J, Sweeting, C. J., Wigham, B. D., Polunin, N. V. C., Hawkes, J. A., Connelly, D. P., Pearce, D. & Linse, K. Linking regional variation of epibiotic bacterial diversity and trophic ecology in a new species of Kiwaidae (Decapoda, Anomura) from East Scotia Ridge (Antarctica) hydrothermal vents. *Microbiologyopen* **4,** 136–150 (2015).