MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

„Genome diversity and free-living lifestyle of chemoautotrophic lucinid symbionts“

verfasst von / submitted by

Bertram Hausl, BSc

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

Master of Science (MSc)

Wien, 2017/ Vienna, 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 830
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Mikrobiologie, Mikrobielle Ökologie und Immunbiologie
Betreut von / Supervisor:	Univ. Prof. Dr. Matthias Horn
Mitbetreut von / Co-Supervisor:	Ass. Prof. Dr. Jillian Petersen

1. Table of content

2. Acknowledgements 5

3. Abstract 6

4. Abstract German 7

5. List of abbreviations 9

6. Introduction 10

6.1 Chemoautotrophic symbioses – providing the food-basis for light-limited environments 10

6.2 The *Loripes orbiculatus* symbiosis 13

6.3 Ecological implications of lucinid symbioses 16

6.4 Aims of this project 17

7. Materials and Methods 19

7.1 Sampling 19

7.2 Screening for the symbiont 20

7.2.1 Testing of FISH probes and optimization of hybridization conditions 20

7.2.2 Optimization of oligonucleotide probes as PCR primers 21

7.2.3 Screening for the symbiont in the environment 21

7.3 Exploration of symbiont phylogeny and microdiversity 22

7.3.1 DNA extraction from gill tissue, PCR amplification and Sanger sequencing 22

7.3.2 Metagenome sequencing and binning 24

7.3.3 Genome analysis and comparison 25

7.4 Sediment community analysis 26

7.4.1 DNA extraction from marine sediment 26

7.4.2 Preparation for amplicon sequencing 26

7.4.3 *In silico* processing and analysis 28

8. Results 29

8.1 Probe binding assays 29

8.2 FISH on pore water filters 30

8.3 Sediment community ... 32
8.4 Symbiont populations phylogeny ... 36
8.5 Host populations phylogeny ... 37
8.6 Metagenome bins ... 38
8.7 Sequence based genome comparison ... 39
8.7.1 Whole genome alignment ... 39
8.7.2 Pan-Genome analysis ... 40
8.8 Metabolic reconstruction and function based genome comparison ... 41
8.8.1 Sulfur metabolism ... 41
8.8.2 Carbon metabolism ... 41
8.8.3 Nitrogen metabolism ... 42
8.8.4 Cellular structures and motility ... 43
8.8.5 Detoxification reactions and other genomic features ... 46
8.9 *Endozoicomonas sp.* and *Shewanella sp.* genomes ... 47
9. Discussion ... 49
9.1 Applicability of FISH-probes for symbiont detection ... 49
9.2 Symbiont and host population phylogeny ... 50
9.3 Variability in the *Cand.* Thiodiazotropha endoloripes genomes ... 52
9.3.1 The *Cand.* Thiodiazotropha pan-genome ... 52
9.3.2 Variation in metabolic capabilities ... 53
9.4 Indications for a free-living lifestyle of the endosymbiont ... 56
9.5 Analysis of the marine sediment ... 58
9.6 Secondary symbionts in *Loripes orbiculatus* ... 60
9.6.1 *Endozoicomonas* ... 61
9.6.2 *Shewanella* ... 61
10. Conclusions and outlook ... 63
10.1 Metabolism ... 63
10.2 Finding the symbiont in the environment ... 64
10.3 Other aspects of the symbiosis ... 65

11. References 67

12. Appendix 77

12.1 Supplementary Materials and Methods 77

12.1.1 DNeasy Blood & Tissue kit protocol 77

12.1.2 Power Soil® DNA Isolation kit protocol 78

12.1.3 QIAquick PCR purification kit protocol 79

12.1.4 ZR-96 DNA Clean-up kit protocol 79

12.1.5 DNA quantification with Quant-iT™PicoGreen® dsDNA reagents kit 80

12.1.6 SOP- Embedding in Steedman's wax for FISH according to Steedman (1957) 81

12.1.7 Fluorescent *in situ* hybridization of gill tissue and filter sections 82

12.2 Supplementary Figures 84

12.2.1 Formamide Series 84

2. Acknowledgements

First, I would like to thank **Jillian Petersen** for giving me the opportunity to do my Master thesis while working in her research group. Thank you for all the supervision, advice and guidance throughout the entire course of my research. I really believe working on these research projects has provided me with a good insight into academic research and the world of microorganisms.

I would like to thank **Julia Polzin** for guiding me throughout my studies and teaching me all the methods necessary to complete my work. Your shared knowledge and experience really helped me getting accommodated with the research environment.

A big thanks to **Ulisse Cardini**, **Sarah Zauner**, **Florian Scharhauser** and the diving team of the field station of the HYDRA Institute for Marine Science on Elba for collecting fresh clam samples, pore water and sediment cores. Likewise, my thanks go to **Julius Simonis** who brought me many clam samples from a research trip to Piran.

Special thanks to **Rebecca Ansorge** and **Stefano Romano** for helping with the bioinformatics, namely the binning and pan-genome analysis.

I want to thank all my fellow colleagues at the Division of Microbial Ecology of the University of Vienna for providing me with a very pleasant work environment and always being there to answer questions, share opinions or assist me in the lab.

Last but not least, thanks to my family which has supported me throughout all my years of learning and is so proud of me that I am almost embarrassed.

3. Abstract

Through their specialized metabolism and carbon fixing capabilities, chemoautotrophic microorganisms provide a source of primary production for a large number of light-limited ecosystems. Ever since their discovery in 1977 chemoautotrophic symbionts have been used as experimental models to answer interesting research questions due to their vast ecophysiological, phylogenetic and habitat diversity. The shallow water lucinid clam *Loripes orbiculatus* (syn. *Loripes lucinalis* & *Loripes lacteus*), living in reducing marine sediments, harbors a sulfur-oxidizing, chemoautotrophic endosymbiont inside its specialized gills.

The present study aimed to establish molecular methods to be able to reliably detect the *Loripes orbiculatus* endosymbiont within the host gills as well as the outside environment and characterize the genomes of distinct symbiont populations.

Fluorescence *in situ* hybridization (FISH) as well as 16S amplicon sequencing were successfully used to screen lucinid gills for the presence of the symbiont, but exhibited some limitations when applied to environmental samples. It was however possible to detect a putative FISH signal of the symbiont during the analysis of sediment pore water collected from the island of Elba, Italy. The analysis of 9 symbiont strains further supported the hypothesis that the symbionts are indeed free-living and more likely to be found in the water column, rather than attached to sediment surfaces. An extensive genome comparison offered a valuable insight into the pan-genome of the symbiont species, revealing an unexpected, large variability in the metabolic capabilities of distinct symbiont strains. During the course of this study, some interesting metabolic pathways, such as the ability to use cyanate as a nitrogen source or C1-compounds as electron acceptors, were uncovered and discussed. The results from this study demonstrate that the bacterial symbionts are specific to their habitat rather than their host species and possess a suit of features that may represent adaptions to different microenvironments. During the binning of the symbiont genomes, additional genomes belonging to the gammaproteobacteria *Endozoicomonas* and *Shewanella* within the gills were uncovered and subsequently characterized through their genomic features. The findings of this study serve to provide a broad insight into the genetic potential of *Loripes orbiculatus* endosymbiont, laying the foundation for further investigations of the unique capabilities of these chemoautotrophic systems.

4. Abstract German

Durch ihren speziellen Stoffwechsel und die Fähigkeit Kohlenstoff zu fixieren, bilden chemoautotrophe Mikroorganismen eine Quelle der Primärproduktion für eine große Zahl an lichtlimitierten Ökosystemen. Seit ihrer Entdeckung 1977 wurden chemoautotrophe Symbionten aufgrund der großen Vielfalt ihrer Ökophysiologie, Phylogenie und Lebensräume, als experimentelle Modelle benutzt um interessante Forschungsfragen zu beantworten. Die Seichtwasser-Muschel *Loripes orbiculatus* (syn. *Loripes lucinalis* & *Loripes lacteus*) lebt in reduzierenden Meeressedimenten und beherbergt Schwefel-oxidierende, chemoautotrophe Endosymbionten in ihren spezialisierten Kiemen.

Ziel dieser Studie war es molekulare Methoden zu etablieren, um den *Loripes orbiculatus* Endosymbionten, sowohl in den Kiemen des Wirts, als auch in der freien Umwelt, verlässlich zu identifizieren und die Genome verschiedener Symbionten-Populationen zu charakterisieren.

Fluoreszenz *in situ* Hybridisierung (FISH), sowie 16S rRNA Amplicon-Sequenzierung wurden erfolgreich angewandt um Kiemen von Luciniden auf die Präsenz des Symbionten zu testen, stießen bei Proben aus der Umwelt jedoch an ihre Grenzen. Es gelang dennoch, während der Analyse von Sediment-Porenwasser von der Insel Elba (Italien), ein potentielles FISH Signal des Symbionten zu entdecken. Die Analyse der Genome von 9 Symbiontenstämmen untermauerte die Hypothese, dass der Symbiont tatsächlich freilebend und eher in der Wassersäule zu finden ist, als angeheftet an der Sedimentoberfläche. Ein genauerer Vergleich der Genome bot einen wertvollen Einblick in das Pan-Genom dieser Symbiontenspezies und zeigte eine unerwartet große Vielfalt an Stoffwechselwegen in den einzelnen Symbiontenstämmen. Im Zuge dieser Studien wurden einige interessante Stoffwechselwege, darunter die Verwendung von Cyanaten als Stickstoffquelle oder die Nutzung von C1-Verbindungen als Elektronenakzeptoren im Genom entdeckt und diskutiert. Die Ergebnisse dieser Studie zeigen, dass die Symbiontenstämme eher spezifisch für ihr Habitat sind, als für die Spezies ihres Wirts und eine Reihe von Eigenschaften besitzen, die möglicherweise Anpassungen an unterschiedliche Mikrohabitate repräsentieren. Während des Binnings der Symbiontengenome wurden zwei zusätzliche Genome der Gammaproteobakterien *Endozoicomonas* und *Shewanella* in den Kiemen entdeckt und

ebenfalls durch ihre funktionellen Gene charakterisiert. Die Ergebnisse dieser Studie dienen als weiter Einblick in das genetische Potential des *Loripes orbiculatus* Endosymbionten und legen einen Grundstein für weitere Untersuchungen der einzigartigen Fähigkeiten dieser chemoautotrophen Systeme.

5. List of abbreviations

Abbreviation	Full text
BLAST	Basic Local Alignment Search Tool
Bp	Basepairs
BSA	Bovine Serum Albumin
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAPI	4’,6-Diamidin-2-phenylindol
EDTA	Ethylenediaminetetraacetic acid
ETOH	Ethanol
FISH	Fluorescence *in situ* hybridization
FITC	Fluorescein isothiocyanate
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information (USA)
OTU	Operational taxonomic unit
PBS	Phosphate-buffered saline
PCoA	Principle coordinates analysis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PFA	Paraformaldehyde solution
PQQ	Pyrroloquinoline quinone
RT	Room temperature
SDS	Sodium dodecyl sulfate
SOP	Standard operating procedure
SOX	Sulfur oxidation
TCA	Tricarboxylic acid cycle
TRAP	Tripartite ATP-independent periplasmatic transporters
TRIS	Tris(hydroxymethyl)aminomethane
µm	micrometer
µl	microliter

6. Introduction

6.1 Chemoautotrophic symbioses – providing the food-basis for light-limited environments

Autotrophy, the ability to produce organic compounds by utilizing inorganic substrates, forms the nutrition basis in any ecosystem. For a long time, it was believed that significant primary production could only be achieved through photosynthesis and organisms living in the deep would depend directly on food sources sinking down from surface layers [1]. Thus the sustenance of large animal communities in light deprived deep-sea ecosystems remained puzzling.

It was the discovery and investigation of the gutless and mouthless tubeworm *Riftia pachyptila* in 1977 that changed the view of marine biologists in that regard [2][3]. These marine invertebrates live in large aggregates near hydrothermal vents, several miles deep down in the Pacific Ocean and count amongst the fastest growing animals on earth. A thorough investigation of these worms revealed that they house endosymbiotic bacteria within a specialized organ called trophosome [4]. It soon became apparent that these microorganisms provide an important food source for the worms. To maintain the fast growth rate of their host, the bacteria had to be able to rapidly transport the carbon compounds to the worm, indicating the presence of an alternative autotrophic metabolism. Analysis revealed that the symbionts have the ability to fix inorganic carbon in the form of CO_2 by using reduced sulfur compounds as electron donors (see Figure 1).

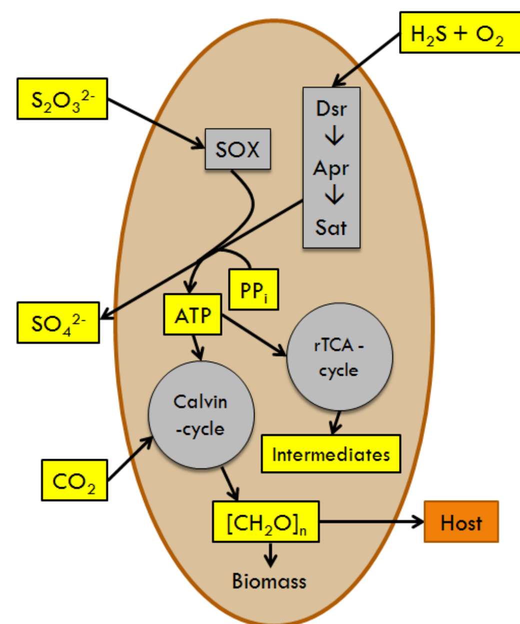


Figure 1: General overview of the carbon dioxide fixation pathways coupled with sulfide-oxidation in a chemoautotrophic symbiont cell. Based on Petersen et al. 2016 [5]

Sulfide oxidation is mediated by a set of three cytoplasmatic enzymes: DsrAB, AprAB and SopT [6]. In addition, many chemoautotrophs can also use the periplasmatic Sox enzyme system to oxidize thiosulfates [7], converting them to sulfite and finally sulfate. The

generated energy in form of ATP and NADH is used in the carbon fixation through the Calvin-Benson-cycle as well as the reductive TCA cycle [7]. The synthesized organic compounds are then either used for the symbionts own metabolism or translocated to the host. This can be achieved through active transport, leaking the metabolites (milking) [8], the digestion of bacteria (farming) or any combination of the three [9][10].

This discovery sparked a new wave of exploration that brought forth an abundance of similar symbioses in a variety of hosts. *Riftia pachyptila* became an experimental system for chemoautotrophic associations, being the first chemoautotrophic system to be intensively studied. The relative simplicity of these systems, involving only two organisms as partners facilitates molecular studies such as isotope composition analyses and allows for easy comparisons of systems.

Meanwhile, many similar systems were discovered in different marine environments ranging from deep-sea habitats, such as whale- and wood-falls as well as cold seeps, to more accessible sites, such as mangrove peats and shallow water coastal sediments [3]. The presence of chemoautotrophic symbioses in shallow water environments has become a blessing for researchers interested in these interactions, as studying shallow water systems removes difficulties and costs incurring with deep-sea studies. Today we know of seven different animal phyla, all of which invertebrates, that host chemoautotrophic symbionts, either inside or on the surface of their bodies [3]. The vast diversity of habitats and associations of symbiotic bacteria is reflected in the huge variety of symbiont groups. While most of these microorganisms are gammaproteobacteria, chemoautotrophic bacteria form a polyphyletic group spread over at least nine known clades [3]. The vast diversity of animal hosts, bacterial symbionts, habitat distributions and metabolic capabilities of these systems are subject of current studies.

The common factor between all these systems is the presence of inorganic energy sources such as sulfur-compounds, hydrogen or methane; substrates generally more commonly found in anaerobic environments. Still, both oxygen and reduced compounds are needed for the symbionts' aerobic metabolisms, providing some challenges for the bacteria. The symbionts of *Riftia pachyptila* [11] and *Lucinoma aequizonata* [12], a clam of the family Lucinidae, have developed the ability to use nitrate as an electron acceptor during periods of oxygen shortage, however this compound is only used as a temporary substitute. Isolated

chemoautotrophic symbionts from *Lucinoma aequizonata* were able to reduce nitrate even under aerobic conditions, although at lower rates [12].

Bridging the oxic and the anoxic layers in the environment is essential but difficult for the bacteria. The association with animals can be seen as an adaptation to solve this problem [13]. A great service provided by the animal hosts is the continuous supply of reduced sulfur compounds or methane as well as oxygen. In order to better fulfill this role and to protect themselves from the toxicity of the sulfide compounds, animal hosts have developed various mechanisms over the course of evolution such as morphological adaptations, migration, storing sulfur in non-toxic form [14] or the production of special sulfur-tolerant hemoglobin [15].

The modes of transmission of the symbiont are different from system to system as well. Bacterial transmission can occur in two major ways. (1) Horizontally transmitted symbionts are either recruited from a free-living population or shared with other organisms living in the same habitat. (2) Vertically transmitted symbionts are directly transferred from the parent to the offspring. These different strategies for transmission have a large impact on the co-evolution of both symbiosis partners. Symbionts with an obligate intracellular lifestyle are vertically transmitted and undergo extensive genome reduction [16]. The transmission represents a genetic bottleneck leading to faster nucleotide substitution rates [17] and a loss of genes which cannot be recovered due to asexual reproduction and genetic isolation of the population (Muller's ratchet) [16][18]. Spanned over a long course of evolution this genetic drift and host restriction can lead to extremely small genomes [16][19]. In extreme cases this continuous genetic degradation can create genomes reaching the currently known lower limit of genome size of 112 kbp found in the insect endosymbiont *Nasuia deltocephalinicola* [20]. Symbionts from an environmental population on the other hand exhibit a great variety in their genomes. The uptake of DNA from other microorganisms in the environment can expand their genome, adding new features which may provide advantages to environmental adaptations [21]. The competition with other bacteria creates a strong selective pressure for advantageous genes making them less likely to be discarded [22].

For a long time chemoautotrophic symbionts were not cultivable, making molecular methods required for classifications and physiological studies [23]. It was only recently that the research group of Distel et al. managed to bring the chemoautotrophic endosymbiont of

the bivalve *Kuphus polythalmia* into pure culture by growing symbiont cells extracted from host gills on selective agar medium plates under microaerobic conditions [24]. Traditionally, symbiont bacteria have been identified through their 16S rRNA or functional gene sequence. Detection methods such as fluorescent *in situ* hybridization (FISH) have allowed for a characterization and localization of target microorganisms directly in the sample. Metagenomics represent another cultivation independent method being used to describe uncultivated microorganisms and their metabolism. This approach allows for the reconstruction of metabolic pathways and helps determining the ecological role of an organism, however it is focused more on individual species and rarely encompasses entire communities [25]. In addition, incubations with radioactive or stable-isotope labeled substrates shed light into the organism's metabolic activity and trophic relationship within a symbiotic association [2].

6.2 The *Loripes orbiculatus* symbiosis

The chosen experimental system for this study is the littoral clam *Loripes orbiculatus* (syn. *Loripes lucinalis* (Lamarck, 1818), *Loripes lacteus* (Poli, 1791)) of the family Lucinidae. It inhabits the sub-oxic zones of fine grained marine shallow water sediments. This group of mollusks provides ideal experimental systems for symbiont studies, as specimens can easily be sampled and kept in aquaria [26]. To date, all examined species in this widespread family of coastal clams form symbioses with chemoautotrophic bacteria [2][27].

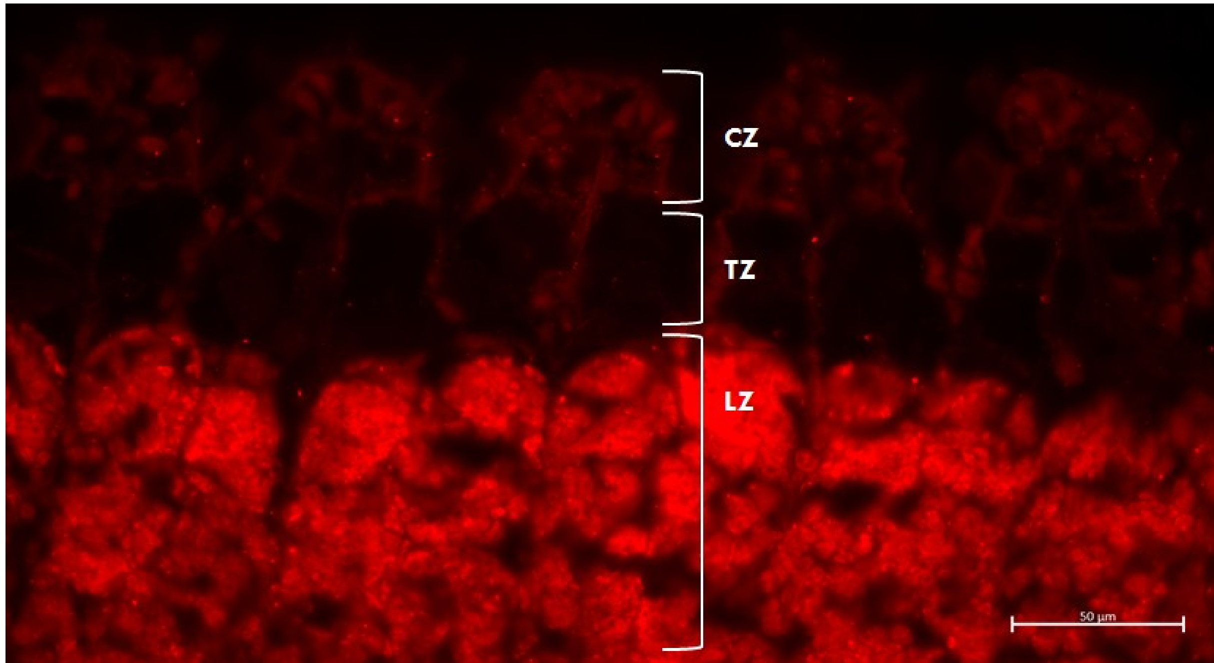


Figure 2: Fluorescence microscopy picture of a gill section of *Loripes orbiculatus* divided into a ciliated zone (CZ), transition zone (TZ) and lateral zone (LZ). The bright fluorescence signals in the lateral zone are sulfur-oxidizing symbiont bacteria.

Over the course of evolution, these bivalves have developed adaptations to accommodate their symbionts. The bacteria are housed inside specialized host-derived vacuoles in the bacteriocytes of the gills [28][29]. The gills themselves have been adapted to provide a suitable environment for the endosymbionts. Their tissue can be divided into three functionally distinct areas: The ciliated zone, the transition zone and the lateral zone (see Figure 2). The ciliates, although reduced, are fully capable of filtering food out of the passing seawater [30][31]. The lateral zone, also referred to as bacteriocyte zone [32], makes up the largest portion of the whole filament volume and is responsible for the thickness and beige color of the gills, due to a large amount of present sulfur inclusions. The entire gill filament is traversed by bacteriocyte channels of different sizes, opening into the ostia of the ciliated zone, allowing seawater to pass through. This provides the bacteriocytes with oxygen and carbon dioxide for the resident bacteria [32]. To reach the oxygenated water, lucinids use their extensible, specialized foot to form a U-shaped tube reaching to the surface layers of the sediment (Figure 3A). Water from the sediment surface is sucked in with the inhalant siphon while waste products are discharged through the exhalant siphon. In addition, reduced sulfide compounds are supplied from the anoxic layer of the sediment column

below by the clams foot. Thus, all the necessary major substrates for the growth of the carbon-fixing endosymbionts and the host itself (H_2S, CO_2 and O_2) are provided by the host. Conversely, the symbionts provide a large part of the bivalve nutrition. A $\delta^{13}C$ value of ~32,7‰ measured in the tissue of *Loripes orbiculatus* indicates a large dependency on metabolites synthesized by their endosymbionts [33]. The degeneration of the digestion system observed in all lucinids [30] further increases the reliance on nutrients derived from the bacteria. In addition to housing symbionts, the bacteriocytes may also fulfill functions similar to lysosomes. Over time the bacteria show clear signs of degradation while the bacteriocytes fill up with granule bodies storing glycogen [34].

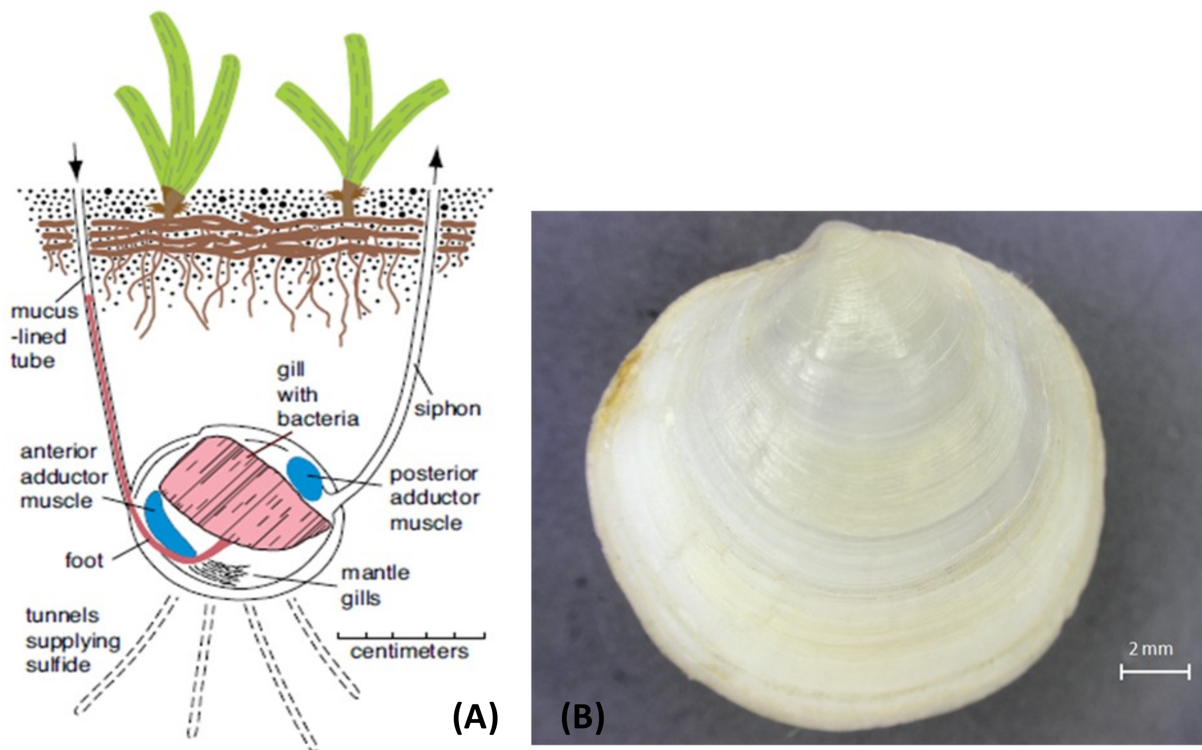


Figure 3: (A) Schematic picture of a lucinid of the genus *Codakia* in its sediment environment. The U-shaped burrows are dug with an extendable foot. From Stanley, 2014 [35] **(B)** Adult specimen of *Loripes orbiculatus* collected from Elba, size ~ 15 mm. Numerous fine concentric lines clearly indicate growth stages.

The endosymbiont, a yet uncultured, sulfur-oxidizing gammaproteobacterium, has been named *Candidatus* Thiodiazotropha endoloripes [5]. Its closest relatives are other sulfur oxidizing symbionts of marine invertebrates such as *Endoriftia persephone* or the nematode *Olavius algarvensis*, as well as free living environmental bacteria belonging to an unclassified clade within the gammaproteobacteria, such as *Sedimenticola thiotaurini* [5]. When

compared on 16S rRNA level, symbionts of several different lucinids show high similarities forming a monophyletic clade. In 2003, Gros et al. demonstrated that the intracellular lucinid symbiont of *Codakia orbicularis* released into the environment is capable of infecting a new host of a different, but closely related species. In this study 100% identical symbiont 16S rRNA sequences were found in five different lucinids (*Codakia orbicularis, Codakia orbiculata, Codakia pectinella, Divaricella quadrisulcata* and *Lingua pensylvanica*) [27]. The infection of different hosts is a strong indication for a horizontal or environmental transmission mode of the symbiont. Due to the close relatedness and high similarity to these systems, a possible horizontal transmission in the *Loripes orbiculatus* and *Cand.* Thiodiazotropha endoloripes association is proposed. In 2013, Espinosa et al. [36] were unable to extract symbiont sequences from anywhere other than the gills when screening the gill, mantle and gonad tissue of the bivalve. The absence of bacteria in the gonads also suggests a non-vertical transmission mode for the symbionts. A non-vertical transmitted lifestyle would require a set of modifications of the symbiont cell as well as corresponding genes. In 2016, Petersen et al. demonstrated that the symbiont is capable of nitrogen fixation, a rare ability amongst chemoautotrophic bacteria [5][37]. This metabolic pathway can be used as a means to promote growth of the bacterium in the nitrogen limited environment of marine sediments, providing benefits for the host. In addition to their chemoautotrophic lifestyle, the bacteria also encode a complete gene-set for the uptake of acetate, via the acetate transporter ActP and other organic compounds via TRAP transporters. This indicates a capability for heterotrophic growth by using these compounds in the TCA cycle [5].

6.3 Ecological implications of lucinid symbioses

Lucinids are often found in protected shallow-water environments below, or near seagrass meadows. In the cooperation between these bivalves and their intracellular symbionts, seagrasses can be seen as a third partner of the symbiosis, influencing the chemical composition of the sediment and all organisms within [38]. Marine macrophytes provide protection from predators, stabilize the sediment and prevent bioturbation [35]. By enriching the sediment with sulfides and supplying oxygen through the roots, seagrasses also provide optimal substrates for the clams and their symbionts [38]. The symbionts,

through their metabolism, remove toxic sulfur-compounds from the sediment surrounding the clam. This benefits the bivalves as well as the seagrasses, as hydrogen sulfide and other sulfur-compounds are extremely toxic to many plants and animals. The presence of lucinids has been shown to lower sulfide concentrations significantly and enhance growth of marine macrophytes [39]. It is likely that these organisms have propagated each others' spatial expansion throughout the course of evolution [35][38].

The recent discovery of nitrogen fixation in lucinids [5][37] illuminated another possible aspect of the tripartite symbiosis. Not only the host, but also seagrasses growing in the surrounding environment may benefit from symbiont diazotrophy. A study by Reynolds et al. [39] found the concentration of pore water ammonium increased when lucinids were present in the sediment. The supply of bioavailable nitrogen compounds may be a significant factor boosting marine macrophyte growth.

The lucinid – endosymbiont system can produce a large amount of biomass as a potential food source for marine predators. A study by Higgs et al. from 2016 [40] delivered first evidence of lucinids finding their way into the diet of Caribbean spiny lobsters. Around 20% of the ingested biomass stems from the chemoautotrophic primary production of the lucinid endosymbiont as was confirmed by stable isotope analyses. Thus, lucinids are important contributors to the productivity of these ecosystems and local fishing industries.

6.4 Aims of this project

Lucinids represent a suit of intensively studied experimental systems that have provided great insight into marine symbioses and the potentials of chemoautotrophic organisms. Although our knowledge about the diversity [3], phylogenies [5][28] and functional morphologies [29][30][33] of these systems is increasing, there are many aspects of these associations, such as establishment of the symbiosis, host recognition and dynamics of the symbiont populations, that still are not very well understood. Despite having been studied for decades some important pathways such as nitrogen fixation in the symbiont have only been brought to light recently [5], demonstrating a vast potential of knowledge still to be gained from these mutualistic associations.

This study aimed to gain insight into the characteristics of symbiont populations of the *Loripes orbiculatus – Cand.* Thiodiazotropha endoloripes system. The sulfur-oxidizing symbiont has only been observed inside the bacteriocytes of the host gills thus far. As the symbiont is thought to be taken up from the environment, a first attempt to confirm this hypothesis was made. In this study, molecular methods such as small subunit rRNA sequencing and fluorescence *in situ* hybridization were used to identify and visualize the bacteria. FISH – oligonucleotide probes were tested in the system of the lucinid gills, before using them on cells from pore water samples immobilized on polycarbonate filters.

The symbiont diversity in lucinids from different sampling locations was first investigated on the 16S rRNA level through Sanger-sequencing. Four samples, clustering into different phylogenetic groups were chosen for an in-depth analysis by metagenome sequencing and whole genome comparison, teasing apart the differences in metabolic capabilities of distinct symbiont populations.

As a final part of the project amplicon sequencing of the small subunit of the ribosomal RNA was used to detect the symbiont in its proposed habitat, the marine sediment of sea-grass beds and explore the microbial community composition within.

Specifically, the purpose of the present project was to:

1) Establish a workflow that allows for detection of different phylotypes of symbiont bacteria within the lucinid host.
2) Explore the diversity amongst different symbiont populations and characterize the functional plasticity and metabolic potential of the bacteria.
3) Screen for the symbiont in the environment outside the clams and analyze the microbial community found in the habitat of the sub-oxic marine sediment.

Considering the huge variety of systems for chemoautotrophic symbioses there is much left to be investigated before we can fully understand them. This project aims to bridge the gaps between the characteristics of these associations, specifically the genetic diversity within the group of sulfur-oxidizing symbionts.

7. Materials and Methods

7.1 Sampling

Samples of live *Loripes orbiculatus* clams were collected near a *Posidonia oceania* sea-grass meadow roughly 400 meters off the coast in a depth of 7 m in the bay of Fetovaia on the island of Elba, Italy. Clams of sizes ranging from 0,5 – 2 cm were found up to 30 – 50 cm buried in the sediment and were dug out by hand during scuba diving. The live animals were kept in an aquarium filled with oxygenated seawater and sediment until further processing. Both demibranches of the gills of chosen individuals were removed and fixed in 4% PFA in 0.01 M PBS with 10% sucrose (pH 7.4). After 12 h the samples were washed in 0.01 M PBS with 10% sucrose three times and dehydrated in a series of ethanol of 30%, 50% and 70% concentration. All sample tubes were sealed, stored at 4°C and delivered to the laboratory in Austria.

Pore water was collected in 30 ml syringes in 5 cm intervals from the sediment surface up to a depth of 60 cm near the clams collection site. The water samples were first filtered through a 20 µm mesh to remove any larger particles. The pre-filtered water was then transferred onto a 0.2 µm hydrophilic isopore polycarbonate membrane filters with a diameter of 25 mm (Merck KGaA, Darmstadt, GER) with a vacuum pump. The polycarbonate filters were then fixed in 4% PFA in 0.01 M PBS with 10% sucrose (pH 7.4). After 12 h the samples were washed three times in PBS with 10% sucrose and dehydrated in a series of ethanol of 30%, 50% and 70% concentration.

Three sediment cores in a depth from 0 – 60 cm (Figure 4B) were taken at the *Loripes orbiculatus* collection site.

Additional clam samples for a biogeography and diversity study were collected from the intertidal zone on the coast of the National Parc Banc d'Arguin in Mauritania and a sea-grass bank 10 m off the coast of Piran, Slovenia in the gulf of Trieste (Figure 4A).

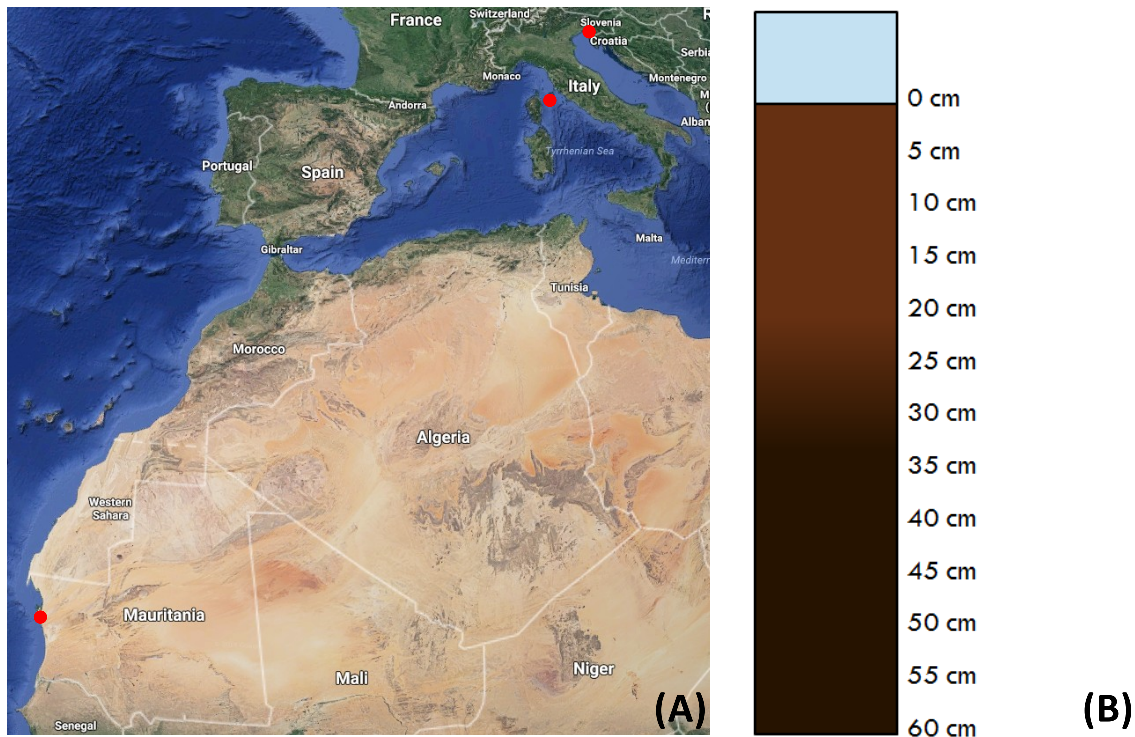


Figure 4: (A) Locations of all three *Loripes orbiculatus* collection sites: Elba(Italy), Piran (Slovenia) and Banc d'Arguin (Mauritania). **(B)** Schematic picture of a sediment core with depth intervals of sampling. In this fine-grained sediment, *Loripes orbiculatus* clams are found in a depth of up to 50 cm.

7.2 Screening for the symbiont

7.2.1 Testing of FISH probes and optimization of hybridization conditions

Two oligonucleotide FISH probes 462-Cy3 and 845-Cy3 (see Table 1) were tested for suitability for symbiont 16S rRNA detection. The optimization of hybridization conditions was carried out on gill tissue of fresh *Loripes orbiculatus* clams. Gills fixed for FISH were put in small embedding cassettes (Simport, Beloeil, QC, Canada) and embedded in Steedman's wax [41]. The block of wax was cut with a microtome into 10 µm thick slices, which were moistened with distilled water and mounted on a superfrost-ultra-plus microscope slide (Thermo Fisher Scientific, Waltham, CA, USA). Before staining the sample was de-waxed and dehydrated in an ethanol series of 90% (3x), 80%, 70% and 50% for 5 minutes each. After drying, FISH was performed on the tissue gill sections. A detailed protocol of the FISH procedure can be found in the appendix. A formamide series was performed to find the optimal hybridization conditions. Formamide concentrations in the hybridization buffer ranging from 0% - 70% in 5% steps were tested with both endosymbiont probes. Additionally,

gill sections were stained with a probe specific for all gammaproteobacteria (Gam42a-Fluos, see Table 1) which served as a positive control. On each microscope slide one gill section was hybridized only with a Non-338-Cy3 probe serving as negative control. All gill sections were additionally stained with the nucleic acid stain DAPI. The stained sections were then examined with an epifluorescence microscope.

7.2.2 Optimization of oligonucleotide probes as PCR primers

16S rRNA amplification and sequencing were used as additional screening methods to detect the symbiont in its environment. Primer annealing specificity of the FISH probes 462-Cy3 and 845-Cy3 was checked by temperature gradient PCR. A gradient of 45°C – 65°C was applied as annealing temperature. Otherwise the same cycling program and reagents as with the regular 16S rRNA were used (see Table 2). The FISH probes served as reverse primers while a general prokaryotic primer 616V was used as forward primer. For better visual evaluation, the DNA template, a gill tissue extraction, was diluted to a concentration below 1 ng/µl. The products were checked by electrophoresis on a 1% Gelred™ (Biotium Inc., Fremont, CA, USA) -stained agarose gel with a GeneRuler™ 1kb DNA Ladder (Thermo Fisher Scientific, Waltham, CA, USA) serving as marker for product size determination. The optimal annealing temperature of the primers that yielded the most PCR products was determined by observing band intensity on the gel.

7.2.3 Screening for the symbiont in the environment

After establishing a working method to detect symbionts in *Loripes* gills, the FISH probes were used on samples from the environment. FISH was performed on filter sections using the oligonucleotide probe 845-Cy3 similar to the gill tissue sections. Probes Gam42a-Fluos, targeting all gammaproteobacteria and its competitor probe Bet42a, targeting all betaproteobacteria (Table 1) were used to provide a positive control, while probe Non-338-Cy3 served as negative control on separate filter sections. DAPI staining was used to distinguish between the auto-fluorescence of any debris and signals from prokaryotic cells.

Despite the pre-filtering, filters with water samples from deeper layers below 40 cm were loaded with dark sediment making signal detection difficult.

Name	Sequence (5' - 3')	Target	Specificity	Reference
616V	AGAGTTTGATYMTGGCTC	16S rRNA	Prokaryotes	Müller et al. (2000) [42]
1492R	GGYTACCTTGTTACGACTT	16S rRNA	Prokaryotes	Brandl et al. (2001) [43]
82F	GAAACTGCGAATGGCTC	18S rRNA	Eukaryotes	Brown&Wolfe (2006) [44]
Medlin B	TGATCCTTCTGCAGGTTCACCTAC	18S rRNA	Eukaryotes	Medlin et al. (1988) [45]
341F	CCTACGGGNGGCWGCAG	16S rRNA	Prokaryotes	Herlemann et al. (2011) [46]
785R	GACTACHVGGGTATCTAATCC	16S rRNA	Prokaryotes	Herlemann et al. (2011) [46]
462-Cy3	TCAAGACCCAAGGTTATTAAC	16S rRNA	*Cand.* Thiodiazotropha endoloripes	This study
845-Cy3	TTAGCTGCGCCACTAAACCCT	16S rRNA	*Cand* .Thiodiazotropha endoloripes	This study
Gam42a - Fluos	GCCTTCCCACATCGTTT	16S rRNA	γ - proteobacteria	Manz et al. (1992) [47]
Bet42a	GCCTTCCCACTTCGTTT	16S rRNA	β - proteobacteria	Alm et al. (1996) [123]
Non-338 – Cy3	ACTCCTACGGGAGGCAGC	none	Negative control	Wallner et al. (1993) [48]

Table 1: Used PCR primers (red), Miseq PCR primers (blue) and FISH oligo-nucleotide probes (orange). All PCR primers were synthesized by Thermo Fisher Scientific. The oligonucleotide probes for FISH were synthesized by biomers.net GmbH.

7.3 Exploration of symbiont phylogeny and microdiversity

7.3.1 DNA extraction from gill tissue, PCR amplification and Sanger sequencing

A first glance at the diversity amongst the lucinid endosymbionts was provided by 16S rRNA sequencing data. DNA was extracted from gill tissue of 30 different animals, 10 originating from the three different sampling sites Elba, Piran and Mauritania, respectively (Figure 4A). The extractions were performed using a Qiagen DNEasy Blood&Tissue kit, following the

added instructions (see appendix). The DNA was eluted in 50 µl elution buffer instead of 100 µl, to increase the concentration. After ensuring sufficient DNA quality and quantity through Nanodrop N-1000 (Thermo Fisher Scientific, Waltham, CA, USA) measurements, the small subunit RNAs of the symbionts and the hosts were amplified by PCR using general bacterial (616V and 1492R) and eukaryotic (82F and MedlinB) primers (Table 1). For a detailed PCR protocols see Table 2. All PCR products were stained with Gelred™ (Biotium Inc., Fremont, CA, USA), checked on a 1% agarose gel through electrophoresis and purified following a QIAquick PCR purification kit protocol (see appendix). The purified samples, including the used primers were sent to Microsynth Austria GmbH for Sanger cycle sequencing. Sequencing chromatograms were manually checked for errors in the program FinchTV (version 1.4.0) [49] and the raw sequences were improved. Forward and reverse sequences were assembled together forming longer consensus sequences. All obtained sequences were aligned against NCBI nucleotide databases with BLAST [50] confirming that they belonged to *Loripes orbiculatus* and its symbionts. Phylogeny trees of the 16S rRNA sequences were calculated and drawn with the program MEGA6 (version 6.06) [51].

16S PCR

Reagents (per 25 µl reaction)	Time (in min)
17.525 µl ddH_2O	1 x 94°C for 04:00
2.5 µl DreamTaq Green buffer (10x)	30 x 94°C for 00:30
2.5 µl dNTP mix (2 mM)	30 x 52°C for 00:45
0.1 µl BSA (20 µg/µl)	30 x 72°C for 00:45
0.625 µl Primer 616V (50 pmol/µl)	1 x 72°C for 10:00
0.625 µl Primer 1492R (50 pmol/µl)	1 x 4°C for ∞
0.125 µl Dream Taq Polymerase (5U/µl)	
1 µl purified DNA template	

18S PCR

Reagents (per 25 µl reaction)	Time (in min)
17.525 µl ddH_2O	1 x 94°C for 04:00
2.5 µl DreamTaq Green buffer (10x)	30 x 94°C for 00:30
2.5 µl dNTP mix (2 mM)	30 x 52°C for 00:45
0.1 µl BSA (20 µg/µl)	30 x 72°C for 00:45
0.625 µl Primer 82F (50 pmol/µl)	1 x 72°C for 10:00
0.625 µl Primer MedlinB (50 pmol/µl)	1 x 4°C for ∞
0.125 µl Dream Taq Polymerase (5U/µl)	
1 µl purified DNA template	

Table 2: Reagents and programs for 16S and 18S rRNA PCR amplifications. All reagents were manufactured by Thermo Fisher Scientific, Waltham, CA, USA. For primer sequences see table 1.

7.3.2 Metagenome sequencing and binning

Four DNA extracts (P7, P2, P9 and M22), containing symbiont DNA of different phylotypes, based on the MEGA6 phylogeny analysis (see Figure 10), were chosen and sent for Illumina metagenome sequencing at the Biomedical Sequencing Facility at the Research Center for Molecular Medicine (CeMM) of the Austrian Academy of Sciences. Quality control of the extracts was provided beforehand through Picogreen® measurements (Quant-iT™ PicoGreen® dsDNA assay kit by Thermo Fisher Scientific, Waltham, MA, USA) of the DNA concentrations (for detailed protocol see appendix). To explore the phylogenetic composition of the metagenome libraries, the assembled small subunit rRNA was classified by the program phyloFLASH (version 2.0) [52]. Quality trimming and filtering was performed using bbduk, a part of the bbmap toolset (version 36.19) [53]. The initial assembly provided by SPAdes (version 3.9.0) [54] was binned with gbtools [55] and then newly assembled. The bins were subsequently named "Phy1", "Phy2", "Phy3" and "PhyM". For the genomes Phy1, Phy2 and PhyM standard GC-Coverage binning was sufficient, for genome Phy3 however, differential binning yielded better results. Reads obtained from Phy3 and PhyM formed additional bins corresponding to the genera *Shewanella* and *Endozoicomonas* as identified by phyloFLASH. All contigs shorter than 800 bp were removed from the bins. The program checkM [56], which uses conserved lineage specific genes to assess the completeness of a bin, was used for quality control. The final bins were uploaded and annotated on the RAST server [57] with default parameters and automated frameshift fixing. A first functional comparison of the genomes revealed genome Phy3 to be missing some pathways that the other genomes had (see results). To test whether these genes were really absent or an error in the binning process, reads from the whole metagenomes of Phy3 were mapped to the other three draft genomes with a similarity threshold of 76%. The properly mapped reads were extracted with samtools [58] and then visualized in the Integrative Genomics Viewer(IGV) [59].

7.3.3 Genome analysis and comparison

For a comparative analysis of endosymbionts genomes from different populations, 5 draft genomes from Petersen et al. [5] were included in this study. Their GenBank files were downloaded from the NCBI server under the accession numbers LVJW00000000 (*Cand*. Thiodiazotropha endoloripes A), LVJX00000000 (*Cand*. Thiodiazotropha endoloripes B), LVJY00000000 (*Cand*. Thiodiazotropha endoloripes C), LVJZ00000000 (*Cand*. Thiodiazotropha endoloripes D) and LVKA00000000 (*Cand*. Thiodiazotropha endoloripes E) and all 9 genomes were annotated in RAST. In the subsequent analyses the five added genomes were named “SymA – E”.

The 16S rRNA as well as the cytochrome c oxidase subunit 1 (CO1) gene sequences were extracted from the nine metagenomes Phy1 – 3, PhyM and *L. lucinalis* A– E and used to create phylogenetic trees with the program MEGA6 [51] to explore the symbiont and host phylogeny. The CO1 genes of *Loripes lacteus* (acc. no. NC_01327.1), *Lucinella divaricata* (acc. no. NC_013275.1) as well as the partial CO1 sequence of *Codakia orbicularis* (acc. no. KC429121.1) were obtained from the GenBank database and added to the host phylogeny analysis.

To learn more about the specific traits of the symbionts, the genomes were characterized by a phenotyping software using the PICA (Phenotype Investigation with Classification Algorithms) framework [60]. The current version of the software is available freely at http://kiwi.cs.dal.ca/Software/PICA [60]. PICA offers plug-ins enabling the use of support vector machines (SVM); machine learning algorithms capable of finding the best split between two training sets. The training set used is based on a list of genomes published by Feldbauer et. al. [61]. The genetic features used to characterize intracellular lifestyle are mostly negative predictors i.e. their absence indicates an intracellular lifestyle [61]. The output is a text file with a YES or NO answer to the phenotype in question.

A Pan-genome analysis was performed by clustering orthologous genes with the tool Get_Homologues [62]. The clustering algorithms of OrthoMCL [63] were used to group the genes of all 9 genomes into orthologous gene families (=orthologs). Orthologs are defined as sequences with an amino acid identity of 50% or more along at least 50% of the longest protein sequence. This definition of gene families has been successfully established for

genome comparisons [64][65]. The number of unique, core- and cloud genes was extracted from the resulting list of orthologs. The Get-Homologues tool also offers the possibility for visualization of the analysis using auxiliary scripts.

For a functional comparison of all genomes Pathway-tools (version 20.5) [66] as well as RAST were used. Pathway Tools provides an overview of all metabolic pathways in the cell and can highlight differences between two or more genomes while RAST only offers a pairwise functional comparison. Interesting pathways were examined individually.

7.4 Sediment community analysis

7.4.1 DNA extraction from marine sediment

To conduct an analysis of the microbial community within the hypoxic coastal sediment, the habitat of *Loripes orbiculatus*, fragments of 16S rRNA sequences from extracted sediment DNA were obtained through high throughput Miseq-Illumina multiplexed amplicon sequencing. Total DNA was extracted from the sediment with a PowerSoil® DNA isolation kit (MO BIO Laboratories Inc, Carlsbad, CA, USA) following the provided protocol (see appendix). The DNA concentration was measured with Nanodrop N-1000 revealing a significant decrease of DNA concentrations with increasing sampling depth. This may be due to an extraction bias caused by humic substances or other inhibitory compounds, or reflect a decrease in organism abundance in deeper layers. Aliquots of the extracts were normalized to a concentration of ~ 4 ng/µl DNA by diluting before they were used for further steps.

7.4.2 Preparation for amplicon sequencing

To prepare the samples, two PCR steps were performed, according to an SOP based on Herbold et al., 2015 [67]. In the first step PCR, the headed-primer set 341F and 785R (Table 1), which targets the small subunit rRNA of all prokaryotes, producing amplicons with a length of 514 bp, was used. This PCR was performed in triplicates (25 µl reactions) with 25 cycles using 1 µl of DNA, extracted from different layers of sediment cores, as template.

Additionally, three mock communities and one negative sample containing no DNA were prepared for sequencing, for normalization and as control. The products of the first step PCR contained a head sequence (5'- GCTATGCGCGAGCTGC -3') at their ends which is the target binding site for the barcodes in the second step PCR. The triplicates were pooled and purified using a ZR-96 DNA cleanup-kit™ (Zymo Research, Irvine, CA, USA).

In a 50 µl reaction with 5 PCR cycles, using 3 µl of the purified product from the first step PCR as template, the barcodes added another 8 bp on both ends of the amplicons used to identify the amplicons in a pooled sample. For detailed parameters of all PCR steps see Table 3. The final barcoded products were checked on an agarose gel, purified using the ZR-96 DNA cleanup-kit™ and quantified by Picogreen® measurements. All samples were pooled together in equimolar amounts containing about 20 x 10^9 amplicon copies and sent for sequencing at Microsynth Austria.

1st Step PCR for MiSeq sequencing

Reagents (per 25 µl reaction)	Time (in min)
18.025 µl ddH_2O	1 x 94°C for 04:00
2.5 µl DreamTaq Green buffer (10x)	25 x 94°C for 00:30
2.5 µl dNTP mix (2 mM)	25 x 52°C for 00:45
0.1 µl BSA (20 µg/µl)	25 x 72°C for 00:45
0.625 µl Primer 341 F (10 µM)	1 x 72°C for 10:00
0.625 µl Primer 785 R (10 µM)	1 x 4°C for ∞
0.125 µl Dream Taq Polymerase (5U/µl)	
1 µl purified DNA template	

Barcode PCR for MiSeq sequencing

Reagents (per 50 µl reaction)	Time (in min)
36.15 µl ddH_2O	1 x 94°C for 04:00
5 µl DreamTaq Green buffer (10x)	5 x 94°C for 00:30
5 µl dNTP mix (2mM)	5 x 52°C for 00:30
0.2 µl BSA (20 µg/µl)	5 x 72°C for 00:45
0.8 µl Barcode (50 µM)	1 x 72°C for 10:00
0.25 µl DreamTaq Polymerase (5U/µl)	1 x 4°C for ∞
3 µl purified template DNA	

Table 3: PCR protocols for MiSeq sequencing of marine sediment communities. All reagents were manufactured by ThermoFisher Scientific, Waltham, CA, USA. For primer sequences see table 1.

7.4.3 *In silico* processing and analysis

Obtained raw Miseq reads were trimmed, filtered and assembled discarding reads with less than 75% of the expected read size and removing the barcode primer sequences. The processed sequences with a length of roughly 400 bp were classified by the RDP classifier of the mothur software [68]. OTUs were defined as sequences which had 97% similarity with each other. All obtained OUT sequences were aligned to a full length 16S rRNA sequence of the lucinid symbiont with BLAST. Statistical analyses of the microbial communities in all samples were done in R [69] using the packages "vegan" [70] and "phyloseq" [71]. The results were plotted with the graphics package "ggplot2" [72]. Only the 30 most abundant OTUs were used to display the community composition in each respective sample for a better overview. The beta diversity between all samples was calculated using the Bray Curtis dissimilarity which was plotted in a principle coordinates analysis (PCoA).

8. Results

8.1 Probe binding assays

Probe 845-Cy3 was successfully applied for labeling 16S rRNA of symbiont cells within the gill tissue. Both, probe 845-Cy3 and Gam42a-Fluos were able to bind well using a hybridization buffer with 35% formamide concentration. At this formamide concentration the probes exhibit optimal stringency, allowing for a combined use of both probes and can therefore be used for symbiont detection.

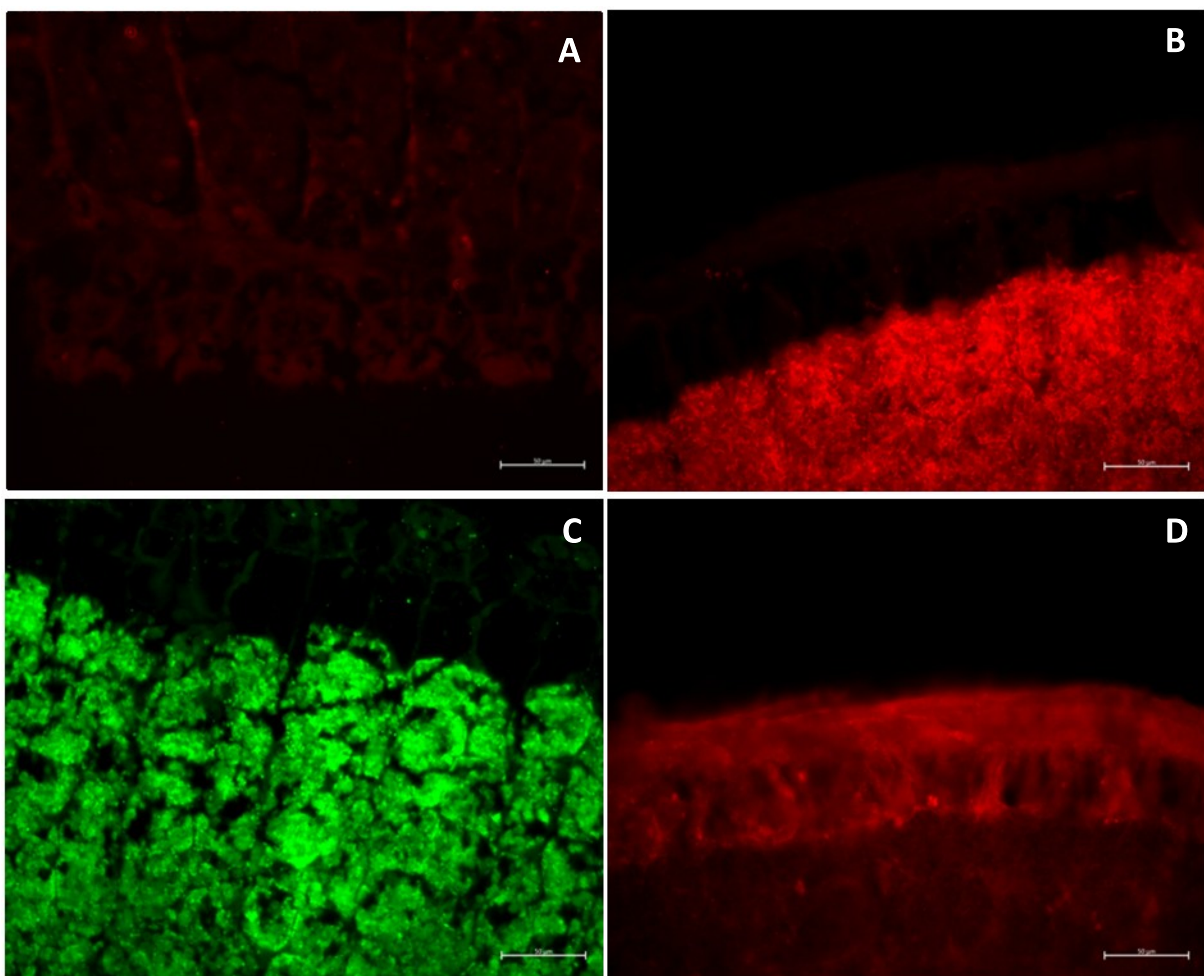


Figure 5: Fluorescence microscopy pictures of *Loripes orbiculatus* gill sections, each labeled with a different FISH-probe. Red channel: Rhodamine B (emission: 575 – 640nm) , Green channel: FITC (emission: 515 – 550nm). Scale bars: 50 µm. Formamide concentrations in hybridization buffer: 10%.
(**A)** 462-Cy3 probe; **(B)** 845-Cy3 probe; **(C)** Gam42a-Fluos probe; **(D)** Non-338-Cy3 probe

Despite matching to its target sequences perfectly, probe 462-Cy3 did not bind in any cells in the gill sections at any formamide concentration in the hybridization buffer. There was no difference in signal intensity between the gills labeled with the 462-Cy3 probe and the negative control (see Figure 5A & D). Any brighter dots in Figure 5A are not signals but represent increased auto-fluorescence caused by inclusions in the tissue, non-hybridized probe remains that were not completely washed away or unspecific probe binding.

8.2 FISH on pore water filters

During microscopy examination of fluorescence labeled pore water samples signals from the Cy3 – labeled probe 845 were observed (Figure 6A – D). In a multiple probe approach, target bacteria were labeled with the nucleic acid stain DAPI, a FITC labeled Gam42a-Fluos probe targeting gammaproteobacteria as well as a Cy3-labeled 845 probe targeting *Cand.* Thiodiazotropha endoloripes specifically (Table 1). The hierarchical approach of multiple labeling gives increased confidence of the signal, decreasing the chance of false positives. Cells labeled with all three labels were present in low abundances on the filters and could only be observed in pore water samples from the upper layers up to 10 cm depth. One triple labeled cell can be seen in the center of Figure 6D. The diameter of the signal seems to correspond with described symbiont sizes of around 2 µm and rod-shaped morphology [34]. In general, cells labeled with all three dyes occurred only in rather low abundance representing an estimated proportion of less than 1% of all cells. In deeper layers increasing amounts of sediment and debris particles were obstructing the view, while causing a strong background of auto-fluorescence, impeding signal detection.

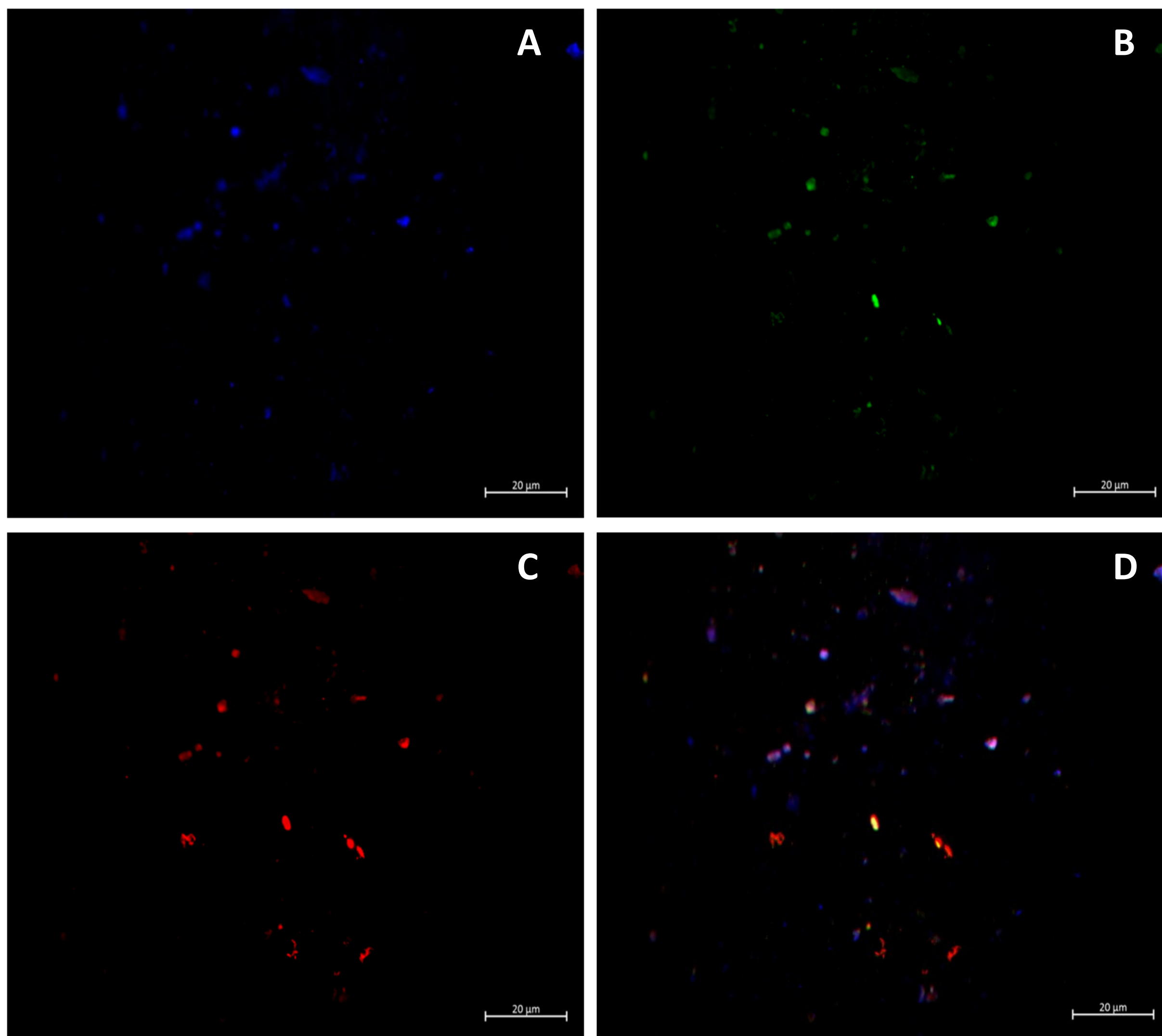


Figure 6: FISH preparations of pore water from the sediment surface fixed on a polycarbonate filter. All 4 pictures represent the same section on the same filter only different fluorescence channels. Scalebars: 20 µm. **(A)**: DAPI staining. **(B):** Gam 42a probe **(C):** 845-Cy3 probe. **(D)** Overlay of all 3 channels. Cells stained with all three dyes adopt a yellow color.

8.3 Sediment community

Classification of the 16S rRNA sequences resulted in a grouping into 763 OTUs across all 12 sediment-layer samples from 0 – 55cm. Some of the most abundant OTUs in the samples were initially classified as chloroplasts. A BLAST search determined that they most likely stemmed from diatoms and foraminifera but it cannot be excluded that they may originate from seagrass debris which was very abundant in the sample. As this analysis is focused on the prokaryotic community these OUTs were excluded from all further analyses. The remaining OTUs were identified as bacteria known from marine sediment environments. When the obtained 16S rRNA amplicon sequences were aligned against the full length 16S rRNA sequence of the *Loripes orbiculatus* symbiont the symbiont seemed to be absent in all sediment samples with the next closest related OTU, an unclassified gammaproteobacterium, being 95% identical in its 16S rRNA sequence.

The analysis of the thirty most abundant OTUs, serving as a proxy for the overall community in the marine sediment (Figure 7A & B), exhibited a large variability between samples. Only the "rare biosphere", making up less than 25% of the total community was excluded from the analysis to make it more concise. On the phylum level, the microbial community is dominated by Proteobacteria which constitute up to two thirds of all OTUs. Their abundance starts to decrease at a depth of 25 cm, instead the appearance of Acidobacteria and a vast increase in Chloroflexi abundance can be observed starting at that depth.

The order of Xanthomonadales exhibits a steady abundance decrease with increasing depth, being completely absent in the deepest layer. They seem to be replaced by a rising amount of vadinBA26 and GIF3, both of which belonging to the Chloroflexi phylum.

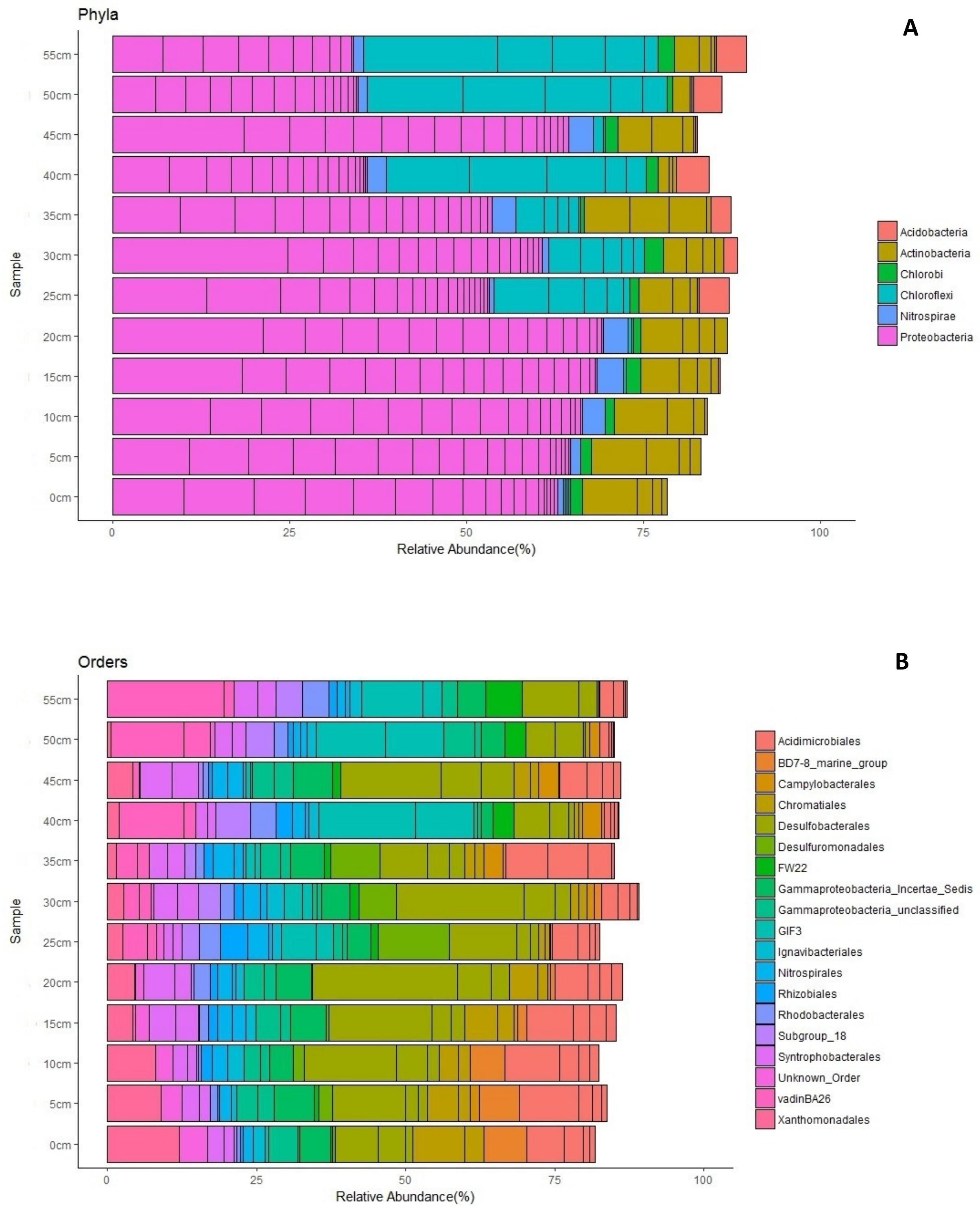


Figure 7: Community compositions in different depth layers of the coastal sediment, showing the relative abundances (%) of dominant taxa on the phylum **(A)** and order level **(B).** Only the 30 most abundant OTUs, based on 16S rRNA amplicon sequencing are plotted.

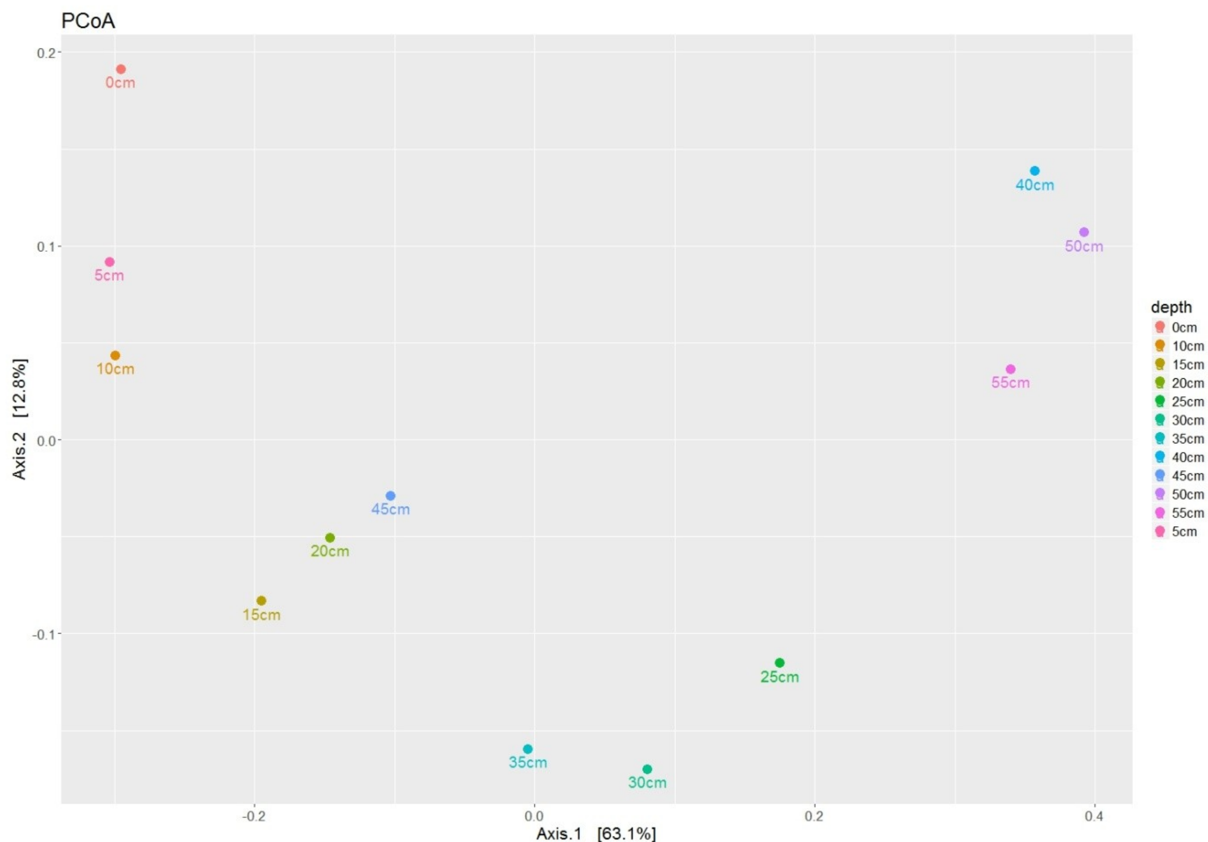


Figure 8: Principal Coordinates Analysis depicting the Bray-Curtis dissimilarity as a means to measure the beta diversity between the sediment samples of different depths.

The shift in the microbial sediment community between depth layers is illustrated in the principle coordinates analysis of the Bray-Curtis dissimilarity in Figure 8. The microbial communities can roughly be grouped in an upper layer community (0 – 20 cm), an intermediary community (25 – 35 cm) and a deep layer community (40 – 55 cm). The 45 cm sample remains as an outlier, exhibiting a composition closer to the upper layer communities.

For a more detailed insight into the different community compositions at different layer depths the 20 most abundant OTUs have been analyzed more closely. The heatmap in Figure 9 depicts these OTUs showing their presence or absence in different sediment depths causing the disparity between samples. The patterns reveal a low abundance of vadinBA26, GIF3 and Subgroup 18 OTUs, all of which belonging to the Chloroflexi phylum, in the upper sediment layers. The steady decrease of Xanthomonadales as well as Acidimicrobiales can also be observed with increasing depth.

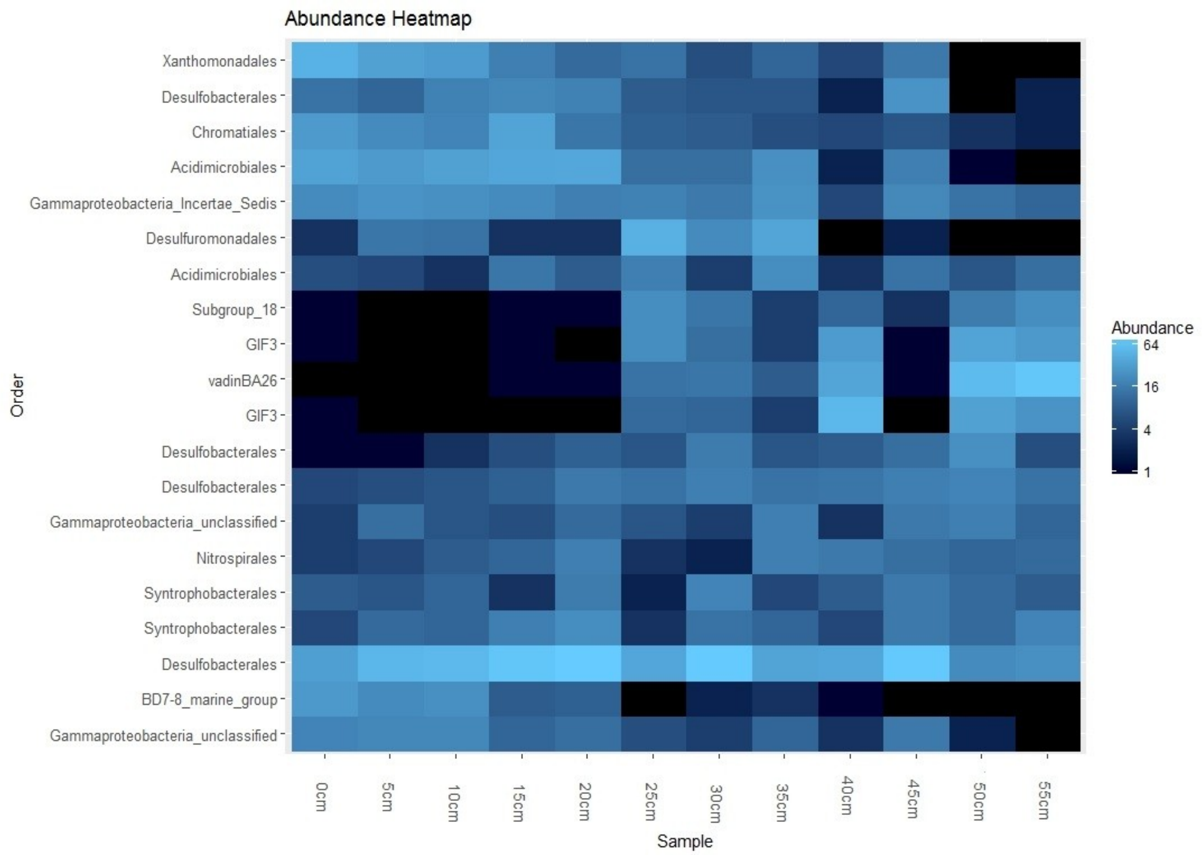


Figure 9: Heatmap depicting the abundances of the 20 overall most abundant OTUs in all sequenced sediment depths. The color gradient ranges from very dark blue to light blue with increasing abundance. OTUs colored in black are completely absent in that sample.

8.4 Symbiont populations phylogeny

The sequences obtained through Sanger-sequencing with a length ranging from ~560 – 1400 bp cluster into four distinct phylotypes. The full length sequence of the ribosomal 16S rRNA of *Cand.* Thiodiazotropha endoloripes is 1492 bp, which was not be completely covered by our chosen sequencing approach. Sequences of small sizes around 800 bp or less, most notably many of the Mauritania sequences, stem from unidirectional sequencing due to sequencing errors where sequencing in one direction failed.

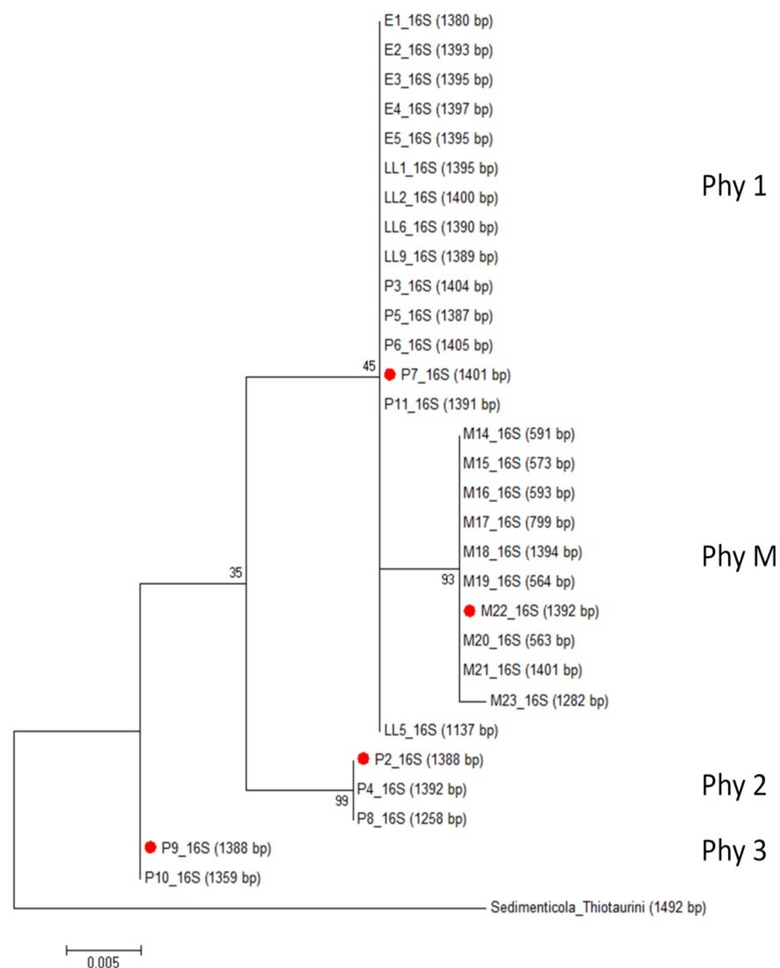


Figure 10: Maximum likelihood phylogeny tree of 30 16S rRNA sequences from *Cand.* Thiodiazotropha endoloripes. The tree was generated by the program MEGA6 with the Kimura 2-parameter model [73] in default settings and assessed by 500 bootstrap replicates. The support values are indicated below the branches (in percent), length of sequences in parentheses. Sequences P2 – P11 originate from Piran, M14 – M23 from Mauritania and LL1 – 5 as well as E1 – E5 were collected from Elba. The scale bar indicates phylogenetic distances. The full length 16S rRNA sequence of *Sedimenticola thiotaurini* was chosen as an outgroup. Samples used for metagenomes sequencing are marked with a red dot. The names of these newly generated genomes are indicated on the right.

Within the four 16S rRNA phylotypes there are little to no nucleotide differences. The nucleotide variation spans from 10 mismatches in the alignment (Phy1/Phy3) to 32 mismatches (Phy2/PhyM). The sequences were later verified with the corresponding 16S rRNA sequences obtained from the metagenomes sequencing which also confirmed the resulting phylogenetic tree and the presence of four distinct phylotypes amongst the samples.

8.5 Host populations phylogeny

Except for a low number of random single nucleotide polymorphisms, 18S rRNA sequences of all hosts were identical. A more suitable phylogenetic marker for eukaryotes was provided by the Cytochrome c oxidase 1 (CO1) gene, the large subunit of the cytochrome c oxidase complex, encoded in the mitochondrial genome with a length of over 1550 bp. These sequences were extracted from the metagenomes and used for the construction of the phylogenetic tree in Figure 11.

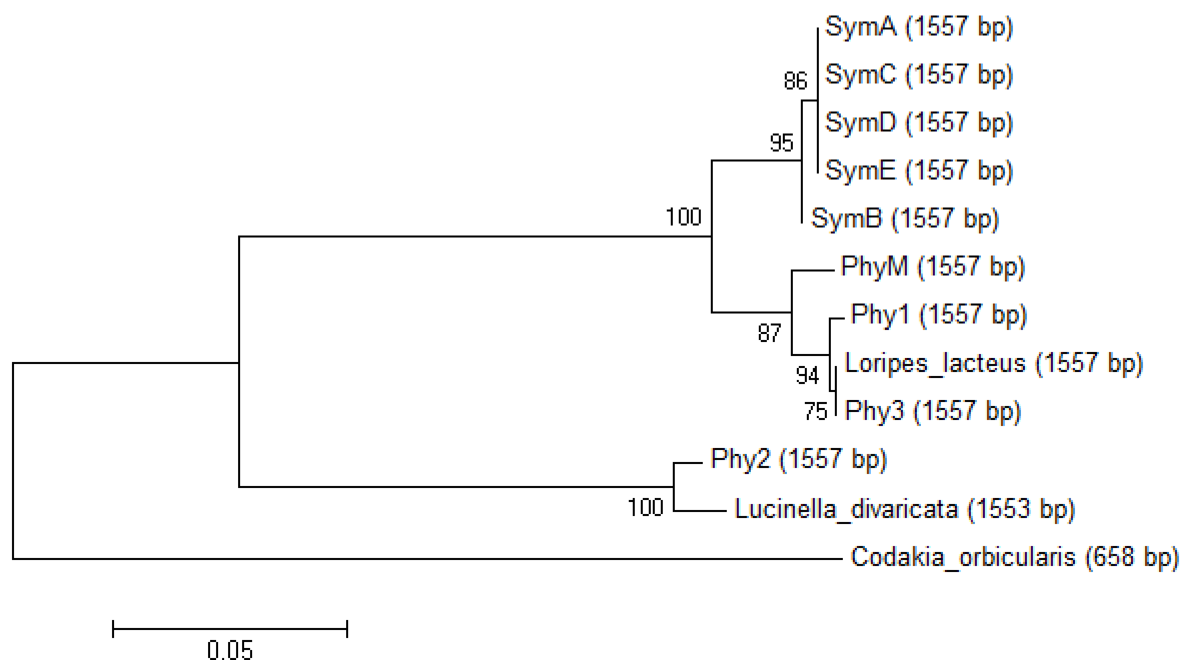


Figure 11: Maximum likelihood phylogeny tree of the mitochondrial cytochrome c oxidase subunit 1 gene of 12 lucinids. The tree was generated by the program MEGA6 with the Kimura 2-parameter model [73] in default settings and assessed by 500 bootstrap replicates. The support values are indicated below the branches (in percent), length of sequences in parentheses.

8.6 Metagenome bins

The phylogenetic composition of the metagenome libraries was analyzed based on raw 16S rRNA reads. Analysis revealed that the metagenomes were largely dominated by only one bacterium species per clam individual: the gill endosymbiont *Cand.* Thiodiazotropha endoloripes. However, two metagenomes contained a second very abundant 16S rRNA sequence. The DNA extract taken from the *Loripes orbiculatus* individual "Phy3", collected in Piran, delivered reads from both the endosymbiont as well as another gammaproteobacterium classified as belonging to the *Endozoicomonas* genus. The extract from the individual "PhyM", from Mauritania, contained reads belonging to *Shewanella*, another gammaproteobacterium. Both of these additional bins were also assembled and annotated, thus creating six genome bins from four metagenomes. It is noteworthy that the *Shewanella* reads in the "PhyM" metagenome had a very high abundance, reaching the same levels as the reads of the primary symbiont. The number of sequences identified as protein coding sequences in the symbiont genomes ranged from 4088 in the "PhyM" individual to 4755 in the "Phy2" individual. The symbiont genomes are split into 48 (in "Phy1") to 269 (in "Phy3") contigs. The genomes of the other two bacteria *Shewanella* and *Endozoicomonas* had significantly more contigs (750-755). All assembled genomes reached high completeness levels from 93,82% up to 100% based on the presence or absence of marker genes specific for the gammaproteobacteria lineage.

	Phy1	Phy2	Phy3	PhyM	Endozoicomonas	Shewanella
Size (bp)	4,914,068	5,125,948	4,590,613	4,690,205	4,590,613	4,359,926
# contigs	48	172	269	232	755	750
GC -content	51.9%	47.3%	48.0%	52.9%	50.0%	50.4%
N50	287,589	72,961	58,063	100,616	12,872	22,094
est. completeness	100%	99.72%	99.72%	98.88%	93.82%	98.88%
Coding sequences	4,335	4,755	4,640	4,088	4,248	3,727
RNAs	54	50	51	50	73	90

Table 4: Statistics of assembled bins from four metagenomes. Completeness is estimated by CheckM [56], number of coding sequences and RNAs are predicted by the RAST [57] annotation.

8.7 Sequence based genome comparison

8.7.1 Whole genome alignment

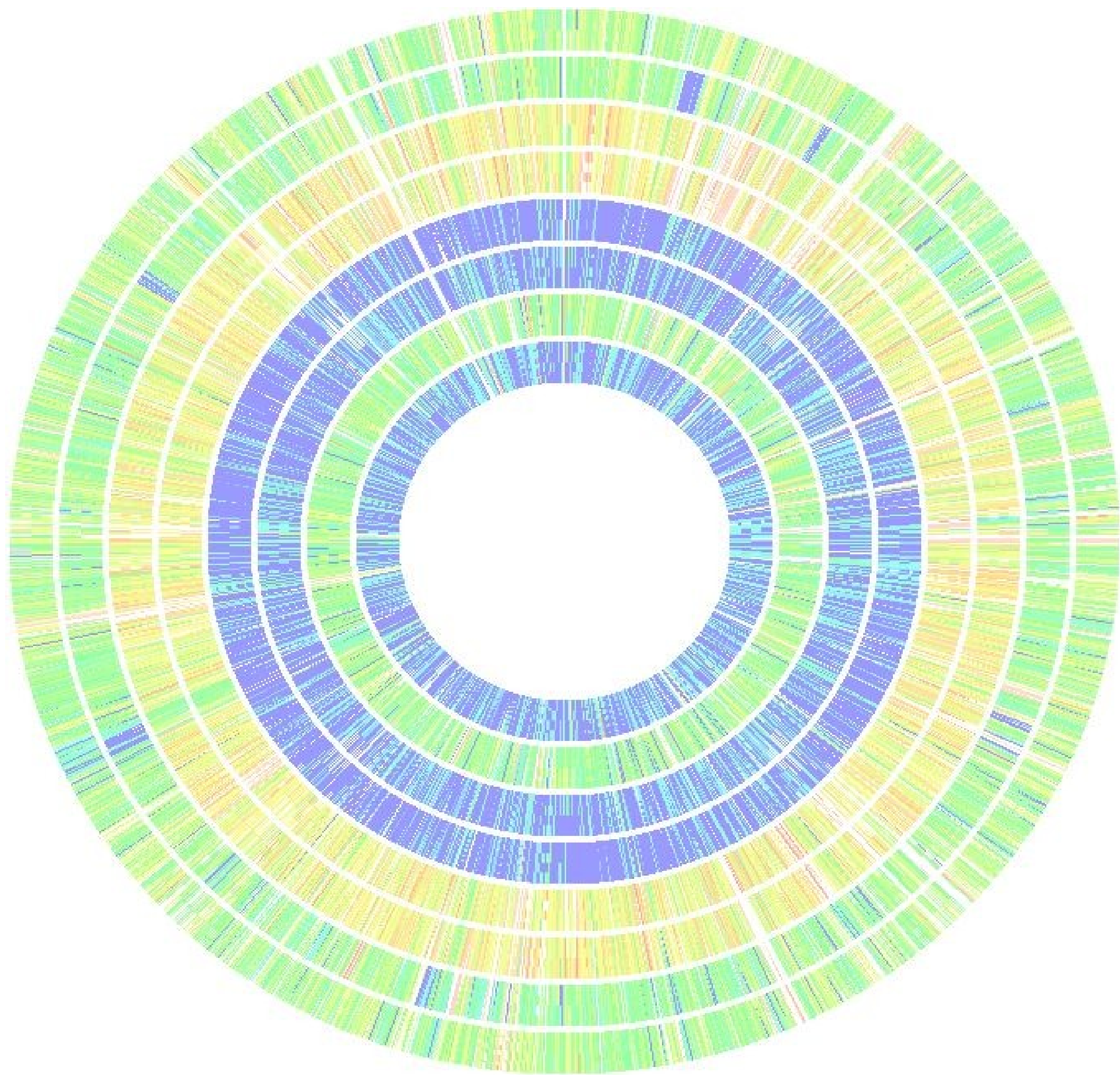


Figure 12: Visualization of a whole genome comparison created with RAST. Comparison is based on unidirectional and bidirectional best BLAST hits of protein sequences. The reference genome is the SymA genome. Every lane corresponds to one genome. Gaps indicate that genes in the SymA genome at that position are not present in the genome of that lane. Genomes from outward – inward: PhyM, Phy1, Phy2, Phy3, SymC, SymD, SymE, SymB.

BLASTp based whole genome alignments revealed very high protein sequence similarities between the Elba symbiont genomes SymA – D. Despite the, less than 0,1%, sequence dissimilarity on the 16S rRNA level, SymE, Phy1 and PhyM exhibit a larger dissimilarity on the whole genome level compared to SymA (Figure 12). The protein sequences of Phy2 and Phy3 differ significantly from those in all other genomes retaining a sequence similarity of roughly 70 – 90% on average compared to SymA. Both of these genomes contain several genomic islands consisting mostly of sequences annotated as "hypothetical proteins".

8.7.2 Pan-Genome analysis

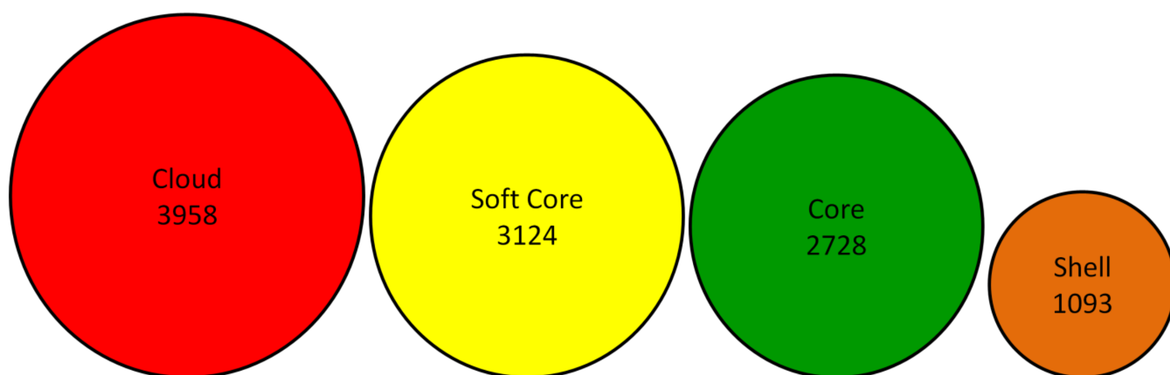


Figure 13: Visual representation of the pan-genome of the nine *Cand.* Thiodiazotropha endoloripes genomes. The sizes of the circles representing the respective parts of the pan-genome are in relation to their number of genes. Definitions: **Cloud genome** = Genes occurring in only 1 or 2 genomes; **Soft Core genome** = genes occurring in 8 or 9 genomes; **Core genome** = genes occurring in all genomes; **Shell genome** = genes occurring in 3 – 7 genomes.

An analysis of the pan-genome revealed a surprising variability between the nine *Cand.* Thiodiazotropha endoloripes genomes. 8175 gene clusters were identified as orthologous groups (50% amino acid sequence similarity along at least 50% of the whole protein sequence) during the analysis. With 2728 genes considered as core genes (present in all compared genomes) about a third of all found gene families are dedicated to the core metabolism of the *Loripes orbiculatus* symbiont. However, the soft-core genes might be a more accurate measure of the number of conserved genes as genes only missing from one genome might be absent due to methods related errors. The remaining 5051 genes (cloud + shell genes) are distributed amongst all nine genomes. Out of these, 3356 genes are unique

to specific symbiont strains. As expected, genomes Phy2 and Phy3 contain the most unique coding sequences with 1042 and 1123 genes respectively, contributing the largest part of the cloud genome. Amongst the unique genes, are several CRISPR and phage related genes as well as mobile genetic elements.

8.8 Metabolic reconstruction and function based genome comparison

The annotation provided by RAST allowed for the construction of an overview of the symbionts metabolism (Figure 14). As expected all 9 symbiont genomes include a gene set ensuring full autotrophic and sulfur-oxidizing capabilities, which are discussed in the following sub-sections.

8.8.1 Sulfur metabolism

All symbionts possess suitable transporters for both the uptake and excretion of sulfate or thiosulfate. The Sox enzyme system is fully functional with *soxXYZAB* being present. The symbionts however do not possess the *soxCD* genes. The entire pathway to oxidize sulfur compounds to sulfate seems conserved amongst all symbiont genomes. The complete dissimilatory sulfite reductase (DSR) complexes (DsrMKJOP) including the co-clustering genes *DsrR* and *DsrS* are present alongside the enzymes adenylyl-sulfate reductase (APR) and ATP sulfurylase (SAT).

8.8.2 Carbon metabolism

All symbiont genomes contain the features necessary for the fixation of carbon dioxide via the Calvin-Benson-cycle. The classical pathway, comprised of twelve genes, is not completely present in the genomes. All symbiont genomes lack the two key enzymes sedoheptulose-1,7-bisphosphatase and fructose-1,6-bisphosphatase. This however does not impede the carbon fixation capabilities in chemoautotrophic symbionts [74].

Transporters for the uptake of various organic compounds can be found within all annotated genomes. This includes several ABC-type and TRAP type transporters (DctPQM) responsible for the uptake of several organic substrates such as malate, succinate or fumarate [75]. The genomes also include a permease (ActP) specific for the uptake of acetate. All genomes have a complete gene set for glycolysis and the citric acid cycle. Additionally, all symbionts have the genomic capability of synthesizing glycogen as storage compounds.

8.8.3 Nitrogen metabolism

All symbionts encode the *nifHDK* cluster and associated genes responsible for nitrogen fixation and ammonium production as has been demonstrated in previous studies [5]. A specific transporter for ammonium can be found in all nine genomes providing an additional ammonium source.

The pathway for the dissimilatory reduction of nitrate, producing molecular nitrogen as final product, is complete in only three genomes (Phy1, Phy2 and SymE). While all genomes contain a periplasmatic nitrate reductase(*NapABC*) with associated electron carriers (*NapFGH)*, genomes Phy3, PhyM and SymA-D lack a cytochrome cd 1 nitrite reductase (encoded by *NirS*) to perform the second step in the reaction forming nitrous oxide out of nitrite. Phy3 even lacks the remaining enzymes nitric oxide reductase (encoded by *NorBC*) and nitrous oxide reductase (encoded by *NosZ*) as well as nitrate specific ABC transporters and regulation proteins. The pathway for assimilatory nitrate reduction via an assimilatory nitrate and nitrite reductase (both NAD(P)H dependent) is entirely missing from genome Phy3 but present in all others.

SymE does not possess a urease enzyme (*UreAC*) nor associated proteins (*UreDEFG)* for direct urea cleavage. Instead urea is degraded via an alternative pathway through a carboxylase and a hydrolase which are conserved in all genomes. SymE also possesses less ABC-transporters for urea uptake, lacking the genes *UrtBCDE* that are present in all other symbiont genomes.

Ammonium is also produced during the L-proline synthesis through L-arginine and L-ornithine degradation. Only genome Phy2 possesses the gene *ArcA* encoding arginase which synthesizes L-ornithine out of L-arginine. Genomes Phy2, Phy3 and PhyM are lacking gene

ArcB required for the desamination of L-ornithine forming the proteinogenic amino acid L-proline. However, all genomes possess the genes *proABC* encoding the pathway for another way of L-proline biosynthesis from glutamate.
Genome Phy3 possesses a unique pathway for ammonium production being the only genome containing genes for the uptake and degradation of cyanate. The resulting carbamate spontaneously forms carbon dioxide and ammonium.

8.8.4 Cellular structures and motility

Overall larger cellular structures seem to be conserved in the symbiont genomes. Each genome encodes a set of 11 genes (*PilBEFJMNOPQZ* and a *FimV* related pilus assembly transmembrane protein) for forming a type IV pilus. Type I, type II and type VI secretion systems have been predicted to be present in all symbionts. A prokaryotic flagellum including motor rotation proteins (*MotAB*) in three copies and other associated genes (*CheABRVYZ*) can be found within all nine genomes. Additionally, methyl accepting enzmyes are conserved amongst all the genomes, some of which are predicted to be aerotaxis sensor receptor proteins(*Aer*).

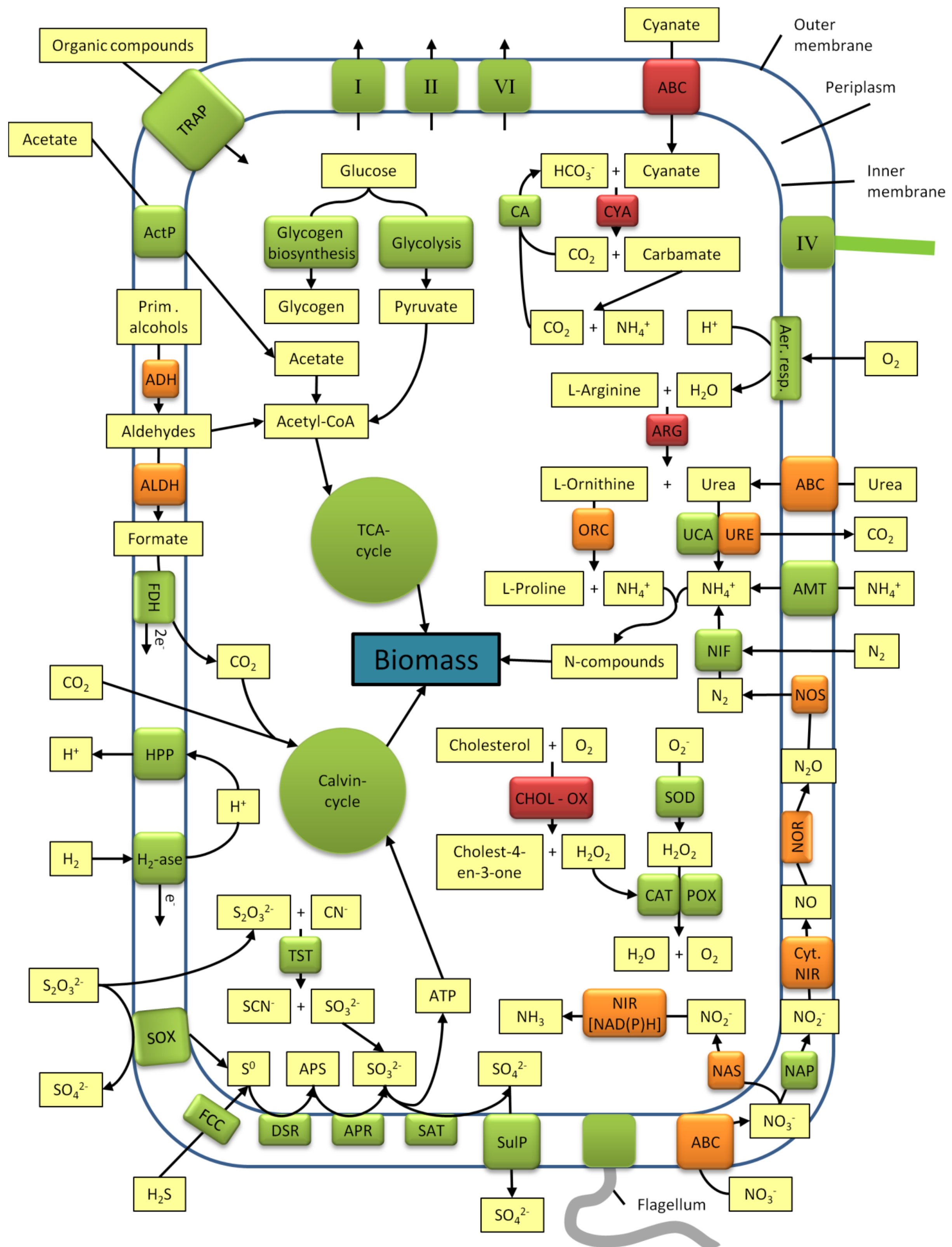


Figure 14: Metabolic model of *Cand.* Thiodiazotropha endoloripes based on a functional comparison of all nine available genomes. Compounds are shown in yellow boxes. Genomic features are represented by green (present in all nine genomes), orange (present in some, but not all genomes) and red (unique to one genome) shapes.

Subsystem	Genetic feature	Abbreviation in fig 14	Presence
Sulfur metabolism	Adenylyl sulfate reductase	APR	All genomes
	Dissimilatory sulfite reductase	DSR	All genomes
	Flavocytochrome c sulfide dehydrogenase	FCC	All genomes
	Sulfate adenylyltransferase	SAT	All genomes
	Sulfur oxidation system	SOX	All genomes
	Sulfate permease	SulP	All genomes
	Thiosulfate sulfurtransferase	TST	All genomes
Carbon metabolism	Acetate permease	ActP	All genomes
	Calvin Benson cycle	Calvin cycle	All genomes
	Glycogen biosynthesis	Glycogen biosynthesis	All genomes
	Glycolysis	Glycolysis	All genomes
	TCA cycle	TCA cycle	All genomes
	TRAP transporters for the uptake of organic compounds	TRAP	All genomes
Nitrogen metabolism	ABC transporter for Cyanate uptake	ABC	Only in Phy3
	ABC transporter for Nitrate uptake	ABC	Missing in Phy3
	ABC transporter for Urea uptake	ABC	Missing in SymE
	Ammonium transporter	AMT	All genomes
	Arginase	ARG	Only in Phy 2
	Cyanase	CYA	Only in Phy 3
	Periplasmatic nitrate reductase	NAP	All genomes
	Assimilatory nitrate reductase NAD(P)H	NAR	Missing in Phy3
	Nitrogenase	NIF	All genomes
	Cytochrome cd1 nitrite reductase (NO forming)	Cyt. NIR	Only in Phy1;Phy2 & SymE
	Nitrite reductase [NAD(P)H]	NIR [NAD(P)H]	Missing in Phy3
	Nitric oxide reductase	NOR	Missing in Phy3
	Nitrous oxide reductase	NOS	Missing in Phy3
	Ornithine cyclodeaminase	ORC	Missing in Phy2; Phy3 & PhyM
	Urea carboxylase	UCA	All genomes
	Urease and accessory genes	URE	Missing in SymE

Cellular structures and motility	Type I secretion system	I	All genomes
	Type II secretion system	II	All genomes
	Type VI secretion system	VI	All genomes
	Type IV pilus	IV	All genomes
	Flagellum	Flagellum	All genomes
	Chemotaxis proteins *CheABRVYZ*		All genomes
	Aerotaxis sensor receptor proteins		All genomes
Detoxification and other features	Quinone dependent alcohol dehydrogenase	ADH	Missing in SymA – D
	Methanol utilization control regulatory protein *MoxX*		Missing in SymA – D
	Methanol dehydrogenase (*xoxF* gene)		Only in Phy2 & PhyM
	Carbonic anhydrase	CA	All genomes
	Catalase	CAT	All genomes
	Cholesterol oxidase	CHOL-OX	Only in Phy3
	Formate dehydrogenase	FDH	All genomes
	Uptake dehydrogenase	H_2-ase	All genomes
	Proton-translocating pyrophosphatase	HPP	All genomes
	Aldehyde dehydrogenase (NAD dependent)	ALDH	Only in Phy2 & Phy3
	Peroxidase	POX	All genomes
	Superoxide dismutase	SOD	All genomes
	Aerobic respiration	Aerobic respiration	All genomes

Table 5: Selected genomic features and comparison between the nine symbiont genomes. This table is using the same color-code as Figure 14: Green: present in all genomes; Orange: present in some genomes; Red: present in only one genome.

8.8.5 Detoxification reactions and other genomic features

A pathway unique to the Phy3 genome is the oxidation of cholesterol. In this reaction, toxic hydrogen peroxide is produced. All symbionts have the capability to remove such peroxides as well as radical oxygen species via the superoxide dismutase and catalase/peroxidase pathway producing molecular oxygen.

Variation was also observed in the degradation pathway of alcohols. Genomes Phy1, Phy2, Phy3, PhyM and SymE revealed a periplasmatic alcohol dehydrogenase converting primary alcohols into aldehydes. With the presence of a specific methanol dehydrogenase, genomes Phy2 and PhyM have a potential way of converting methanol into formaldehyde. Only Phy2

and Phy3 encode an aldehyde dehydrogenase capable of oxidizing aldehydes further into an organic acid such as formate. All nine genomes encode a formate dehydrogenase for formate removal.

As shown in a previous study [5] symbiont genomes SymA – E have revealed the presence of an uptake hydrogenase and a proton pumping hydrophosphatase. As expected the genomes Phy1 – 3 and PhyM also have the capability to produce a proton motive force this way.

All symbionts also encode the necessary enzymes for the electron transport chain associated with aerobic respiration; succinate dehydrogenase, NADH-ubiquinone oxidoreductase and a cytochrome c oxidase.

8.9 *Endozoicomonas sp.* and *Shewanella sp.* genomes

The presence of *Endozoicomonas* and *Shewanella* reads in such high abundances was an unexpected result in the metagenomes analysis. The two bacteria were roughly characterized by the genome annotation in RAST. Both *Endozoicomonas* and *Shewanella* were identified as non-obligate intracellular organisms by the PICA software.

Organism	Features
Endozoicomonas sp. & Shewanella sp.	Anaerobic respiration
	Fermentation of pyruvate to acetate/ethanol
	Glyoxylate cycle
	Ammonium transporters
	Nitrite reductase
	Thiosulfate sulfurtransferase
	Rod shape-determining proteins *RodA* & *MreBCD*
	Type I secretion system
	Type II secretion system
	Type IV pilus
	Chemotaxis proteins *CheA* and *CheR*
	Superoxide dismutase
	Catalase

Only *Endozoicomonas sp.*	Fermentation of pyruvate to lactate
	L-phenylalanine production
	ABC transporters for the uptake of multiple proteinogenic amino acids (His, Arg, Lys, Glu, Gln, Asp) as well as non-proteinogenic amino acids
	Phosphotransferase systems for the uptake of glucose, fructose, maltose, mannose, galactose and other sugars
	Oligopeptide transport system permease proteins *OppBCD & F*
	Glycogen synthesis
	Urea ABC transporters *UrtBCDE*
	Urease
	Type III secretion system
	Extracellular matrix proteins *PelABCDEFG*
Only *Shewanella sp.*	Aerobic respiration
	L-tyrosine production
	Polysulfide reductase *PsrABC*
	Type VI secretion system
	Flagellum and flagellar motility genes
	Chemotaxis proteins *CheBWYZDX*
	Fused spore maturation proteins A&B
	Global Two-component regulator *PrrBA (Reg A& B)*
	Peroxidase
	Phage shock proteins *pspABC & E*

Table 6: Genomic features and comparison of the *Endozoicomonas* and *Shewanella* bins.

9. Discussion

9.1 Applicability of FISH-probes for symbiont detection

The oligonucleotide probe 845-Cy3 was successfully used for the detection of *Cand.* Thiodiazotropha endoloripes in the gills of *Loripes orbiculatus*. Applying the probe with a hybridization buffer with a concentration of 35% formamide provided the most effective results for detecting fluorescence signals. Due to the high abundance of the symbionts, the signal can easily be observed in the microscope, even at lower magnifications. However, a magnification of at least 100x is required to be able to distinguish individual cells. The high density of the symbiont populations and the natural auto-fluorescence of the gill tissue can lead to superimpositions making signal detection difficult. On 10 µm thick sections, these interferences are already strong enough to impede the distinction of single cells. This factor should be reduced as much as possible by cutting only 5 µm thin sections during sample preparation.

Probe 462-Cy3 did not produce any detectable signals after hybridization. However, the probes sequence does match perfectly to a segment of the symbionts 16S rRNA sequence. The reason for the malfunction is not known, but it may be caused by an inaccessibility of the target binding site during hybridization. This is a common problem during probe design, as different species have different possible binding sites [76].

Both oligonucleotide probes were able to amplify 16S rDNA when used as PCR primers. However, this application was only possible in the system of the densely-populated gill tissue. Due to the dominance of only one strain of symbionts, even general bacterial 16S rRNA primers delivered PCR products pure enough for sequencing and specific primers for the symbiont were not needed. Amplifications of 16S rDNA from the water column or marine sediment produced with the oligonucleotide probes were too unspecific and yielded nonsensical reads during Sanger-sequencing.

The sequences of both oligonucleotide probes were aligned to the SILVA rRNA database using the online tool probeCheck [77]. Both probes matched several non-target marine organisms such as *Marichromatium sp.*, *Halochromatium sp.* as well as several other uncultivated gammaproteobacteria, including other lucinid symbionts. This reduces the

usability of these probes to the gill tissue only, as the probes would also target other bacteria from the marine environment outside the clam tissue and would thereby deliver false positives. As a result, the signal detected in Figure 6A – D is not a definite proof of a symbiont found in the environment, despite the multi-probe approach. Identifying a cell as a target organism would require further support from a DNA sequence extracted from the potential free-living symbiont cell, which has not been done successfully so far.

9.2 Symbiont and host population phylogeny

The 16S phylogeny displayed in Figure 10 shows the relationship between symbiont bacteria from different clam individuals. The sequences were obtained through sequencing DNA extracted from the whole gill, therefore the entire symbiont population in one animal is represented by only one sequence. The consequence of this is that any other symbiont phylotypes of lower abundance have not been accounted for. During the phyloFLASH analysis of the assembled metagenomes reads all metagenomes exhibited a low strain heterogeneity with one symbiont phylotype dominating the population within one animal. This indicates that the majority of symbionts might be taken up only once and then proliferate to establish a large population within the gills. This finding corresponds well with studies by Gros et al. in 2012 [78] showing that lucinid clams recruit their symbionts from one bacteria population at an early stage, but are still capable of taking up symbionts throughout their lives. A continuous uptake of only one symbiont strain, as was proposed for *Bathymodiolus* deep-sea mussels and their symbiont is also a possible explanation. It is unknown if the non-dominating symbiont strains have been taken up together with the dominating strain and were then outcompeted or if they were taken up at a later point in time when the dominating symbiont population was already established. The physiological implications of such a low strain heterogeneity of symbionts remain yet unclear.

It is noticeable that the phylotypes depicted in Figure 10 do correspond to sampling sites, implying a geographical factor behind the genetic diversification. With at least 98% 16S rRNA sequence identity however, these phylotypes are still well within the threshold of being considered one single species [79].

The mitochondrial marker gene of the cytochrome c oxidase subunit 1 has long been used as an effective tool for DNA barcoding in eukaryotes [80]. In most species, this gene exhibits large interspecific variation while the intraspecific divergences rarely exceed 2% [81]. The host phylogeny analysis visualized in Figure 11 revealed a surprising find: the host mitochondria sequences from the Phy2 metagenome are very distinct from those of the other hosts. The CO1 gene was only 83% similar to the CO1 genes of *Loripes orbiculatus*. Instead the sequence showed a very high similarity of 98% to the CO1 gene of *Lucinella divaricata*. This finding makes it very clear that the Phy2 genome does not stem from *L. orbiculatus* but represents a different species altogether. It is a distinct possibility that Phy2 is actually a *Lucinella divaricata* individual. The remaining CO1 genes had a variation of 2 – 6% between the individuals collected from Elba and those from Piran. While this value is unusually high, it is debatable if this already accounts for the existence of an overlooked species. One study, analyzing the DNA variation of 227 morphospecies of mollusks, found individual cases of over to 20% intraspecific nucleotide substitutions in the species *Triopha catalinae, Dentalium pilsbryi* and *Rhabdus rectius,* however these are not bivalves [82]. In instances of high intraspecific divergence this circumstance is often attributed to geographical separation [80][82], which certainly could be the case here.

The phylogenetic variability was not reflected in the 18S rRNA sequences which exhibited only very few and seemingly random single nucleotide polymorphisms. Therefore, we can conclude that the 18S rRNA is lacking the resolution to distinguish divergence and the CO1 gene should be used for further analysis. The CO1 gene clearly exhibits a much higher rate of evolution and offers a much better resolution in observing relationships within one species.

9.3 Variability in the *Cand.* Thiodiazotropha endoloripes genomes

9.3.1 The *Cand.* Thiodiazotropha pan-genome

The symbiont genomes from Elba bear a remarkable similarity to each other with the exception of the SymE genome (Figure 12). This genome exhibits dissimilarities on the whole genome level, despite having an identical 16S rRNA sequence. Similarly, Phy1 has the same 16S rRNA sequence as the symbionts in individuals collected from Elba despite stemming from a different collection site. The whole genome alignment demonstrates that a higher genetic diversity is hidden underneath the classical 16S phylogenies. Such diversification naturally has a large effect on the functional features of the respective genomes. Jaspers and Overmann showed that phylogenetically very close relatives based on their 16S rRNA can still exhibit a large genetic and phenotypic variability as early as 2004 [83]. In this study 11 strains of *Brevundimonas alba*, a freshwater bacterium showed significant differences in substrate usage and growth. Signs of such microdiversity can also be seen between the analyzed symbiont genomes as even closely related genomes encode different genes providing various important functions which may impact their ecophysiology.

The core genes make up between 58% and 68% of the entire genome of each symbiont analyzed. The core genome usually represents genes of important housekeeping functions for the bacteria which need to be conserved. However, 452 of the 2728 gene families identified as core genes were annotated as hypothetical proteins without known function, leaving a large part of the genomic traits defining this species yet unexplored. The comparison of 9 genomes already revealed a total of 8175 gene families which is roughly twice the number of genes in one individual symbiont genome. The largest contributors to the number of novel gene families are the genomes Phy2 and Phy3 adding over 1000 unique gene families respectively.

The numbers represented in this study only give a glimpse on the entire variability within the genomes of this symbiont species. In pan-genome analyses the number of core gene families decreases and the number of novel genes increases continuously with the addition of new genomes to the analysis [64]. It is possible that adding more symbiont genomes to the analysis could still expand the pan-genome significantly.

9.3.2 Variation in metabolic capabilities

A major goal of this study was the description of the core and variable metabolic capabilities of the chemoautotrophic symbiont. The core metabolism contains the defining physiological traits of the species. The results of the functional genome comparison displayed in Figure 14 as well as Table 5 and their impact on the bacterium's ecophysiology are discussed hereafter. It has to be mentioned that the metabolic model presented in Figure 14 is purely based on genomic data. With the exception of the symbionts carbon fixation, sulfur oxidation and nitrogen fixation capabilities, metabolic traits of the symbiont have yet to be experimentally confirmed.

One of the most prominent ecological functions of the symbiont is the fixation of carbon dioxide using the energy generated from sulfur oxidation and the allocation of synthesized compounds to the host. Naturally, essential genes in the carbon and sulfur metabolism are conserved in all individuals of this species. Curiously the symbionts do not use the classical Calvin-Benson-cycle but a modified version of it. Two important enzymes (sedoheptulose-1,7-bisphosphatase and fructose-1,6-bisphosphatase) were missing from the Calvin-Benson-cycle pathway. These two enzymes have been reported to be frequently missing from the genomes of autotrophic symbionts; for example in the *Olavius algarvensis* [74] and the *Riftia pachyptila* symbiont [84]. In 2007, Newton et al. proposed that the lack of fructose-1,6-bisphosphatase could be bridged by a proton pumping pyrophosphatase in conjunction with a pyrophosphate dependent 6-phosphofructokinase [85] both of which are encoded in the symbiont genomes. Kleiner et al. [74] argued that also the function of the sedoheptulose-1,7-bisphosphatase may be replaced by the same system.

Similarly to the Calvin-Benson-cycle, the SOX system is also not completely encoded in the symbiont genomes. While the essential *soxXYZAB* genes are present, *soxC* and *soxD* are missing. The absence of *soxCD* does not disable the functionality of the SOX system, but leads to a lower yield of electrons (2 mol instead of 8) for the electron transport chain [7]. The SOX system produces sulfur, which is subsequently oxidized to sulfate by a chain of periplasmatic enzymes that are also conserved amongst all nine symbiont genomes. The last step in this chain, representing the reverse process of the dissimilatory sulfate reduction, yields 1 ATP molecule free to be used in other metabolic reactions such as carbon fixation.

A larger variety can be observed in enzymes needed for the utilization of various nitrogen sources. In 2016 Petersen et al. showed that the symbionts not only possess a complete gene set for the fixation of nitrogen from the environment, including the most important functional marker for N-fixation: *nifH*, but also actively use this pathway inside their hosts [5]. One other common source of nitrogen for chemoautotrophic symbionts is the reduction of nitrate [86 – 88]. There are two major pathways for the reduction of nitrate: (1) nitrate reduction to nitrite via an assimilatory nitrate reductase coupled with the subsequent conversion to ammonia through a NADH or NADPH dependent nitrite reductase $(NO_3^- \rightarrow NO_2^- \rightarrow NH_3)$ or (2) reduction to elemental nitrogen in the process commonly referred to as denitrification $(NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2)$. During the denitrification process, nitrogen compounds are used as electron acceptors for anaerobic respiration. The resulting elemental nitrogen is gaseous and less bioavailable [88] until it is taken up and converted by diazotrophs. This anaerobic way of respiration can be temporarily used during periods of low oxygen availability [11][12]. This is of critical importance as chemoautotrophy requires a lot of oxygen [89] and may be depleting oxygen in the gill. This pathway seems commonly spread, but not completely conserved amongst the chemoautotrophic symbionts. Only 3 out of 9 symbiont genomes encode all enzymes needed for the entire denitrification process. The others are at least able to respire nitric oxide (NO) and nitrous oxide (N_2O) with the exception of Phy3.

In general, Nitrate and subsequently nitrite have been shown to be more important as nitrogen sources rather than oxidants [86]. This also seems to be the case in *Cand.* Thiodiazotropha endoloripes, as all but one genome contained a nitrite reductase that produces ammonia available for biosynthesis. Genome Phy3 is the only genome to be lacking this pathway and is therefore the only one which cannot use nitrate directly for ammonia production.

Instead, this genome seems to have acquired a different way of covering its nitrogen needs: The conversion of cyanate. Phy3 is the only analyzed genome encoding a bicarbonate dependent cyanase to convert cyanate into ammonia and carbon-dioxide. Cyanate is a small molecule, occurring in the environment as a byproduct of urea or cyanide degradation, but may also be formed within the cell [90]. The archaeum *Nitrososphaera gargensis* [90] as well as a *Pseudomonas fluorescens* strain [91] haven been reported to grow on cyanate as their

sole source of energy and nitrogen. The activity of cyanase has the additional effect of removing cyanate which is toxic at higher concentrations. To do so the help of cyanate permeases is required [91], which have also been found in the Phy3 genome. Nitrate concentrations are generally very low in the marine environment [92] and may not cover the nitrogen requirements of the symbiont. Thus, the utilization of cyanate may be an adaption to lower nitrate concentrations in the geochemistry of the host's habitat.

Another valuable source of nitrogen can be provided by the recycling of the proteinogenic amino acid arginine. The hydrolysis of arginine produces ornithine as well as one urea molecule which can be degraded into ammonia. The enzyme catalyzing this hydrolysis has been found in the Phy2 genome. Ornithine can be further degraded to proline under anaerobic conditions, dissociating another ammonia molecule [93]. The deamination of ornithine represents an alternative pathway of proline biosynthesis [94]. In total, three molecules of ammonia can be produced by the degradation of arginine to proline, however no genome encoded all enzymes necessary for the complete pathway.

A novel feature that had not been observed in *Cand.* Thiodiazotropha endoloripes genomes before is the potential use of methanol through a PQQ-dependent alcohol dehydrogenase encoded by the *xoxF* gene. This gene is not conserved in the symbionts found in Elba but is present in all strains from Piran and Mauritania. The gene, a homologue of *mxaF* is widespread in coastal marine environments, where it is found in all known methylotrophic prokaryotes and has therefore been used as a functional gene to identify methylotrophic bacteria [95]. Methylotrophy is the capability of metabolizing reduced C1 compounds such as methane, methanol or formate [96]. Facultative methylotrophy is found within many distinct groups of heterotrophic bacteria and may also be widespread in chemoautotrophs [96]. A major source of methanol in marine sediments is the microbial degradation of organic matter. A large part of this organic matter may be provided by the debris of seagrasses which are a vital part of the lucinids' habitat. Not much is known about the precise methanol concentrations in marine sediments as its volatile nature makes measurements difficult [97]. Methanol can be used as an intermediate compound by both methane producers and methanotrophs, therefore playing an important role in geochemical carbon cycles. The oxidation of methanol by methanotrophs produces toxic formaldehyde which can be assimilated via various pathways or further oxidized to formate [98]. Formate

dehydrogenases capable of oxidizing formate are found within all symbiont genomes and provide an additional source of carbon. The product of formate oxidation; CO_2, is fixed through the Calvin-Benson-cycle [96]. Degradation of methanol has also been linked to nitrate reduction, naming methanol as an important electron donor for denitrification in aquatic environments [99]. Much like the enzymes of nitrate reduction, genes for methanol and aldehyde utilization are part of the variable genome of *Cand.* Thiodiazotropha endoloripes. Only genome Phy2 possesses enzymes for the entire oxidation pathway from methanol to formaldehyde, formate and finally carbon dioxide. The other symbiont strains may only be able to use distinct C1 compounds. There is no indication that any of the symbionts can use methane directly as a carbon source.

9.4 Indications for a free-living lifestyle of the endosymbiont

In this study, it was not possible to find definitive proof for the presence of the *Cand.* Thiodiazotropha endoloripes in the environment outside of its clam host. However, genetic analyses delivered strong indications for a potential extracellular lifestyle.

The first clue is the large size of the genomes. With genomes ranging from ~4,450,000 bp to 5,125,948 bp in size, the genomes clearly fall into the spectrum typical for free-living bacteria genomes [100][101]. Obligate intracellular symbionts have been found to have a very small average genome size of 0,5 – 1,5 Mb [101]. These bacterial genomes usually contain a large number of pseudogenes and repetitive elements [16][18][19][101], which were not found in the genomes in this study. With an average of around 4300 genes, a very large part of the symbiont genome is comprised of coding sequences. This leads to the conclusion that the bacterium is a facultative symbiont also capable of living outside the clam. One other explanation for such a large and non-degraded genome would be that the symbiosis has only recently become obligate for the bacterium and the genome reduction has not progressed very far. Fossil records dating back to the Silurian period, however, already exhibit morphological features supporting the hypothesis of ancient lucinid clams possessing chemoautotrophic symbionts [102]. It is not certain but very likely that the symbiosis is obligate for the lucinid host as the bivalves die when the symbiont population in their gills is depleted.

As previously discussed, the phylogeny of the animal hosts revealed that at least the Phy2 individual represents a different lucinid species. This fact has interesting consequences on our view of the transmission of the symbiont. Finding the same symbiont species in distinct clam species suggests that they either recruit their symbionts from similar bacteria populations in the environment or exchange symbionts between the two species. Gros et al. showed that lucinid symbionts of *Codakia orbicularis* can colonize a larger range of hosts [27]. The same may be true for the *Loripes orbiculatus* symbiont. This symbiont species may be more specific to one habitat rather than one species of hosts.

Another indication for a free-living stage is the large genetic variability between the *Cand.* Thiodiazotropha endoloripes genomes. The pan-genome analysis revealed that only 2728 genes are classified as core genes (50% protein sequence identity along 50% of the gene sequence) for all genomes. Conversely this means that between 32 – 42% of the coding sequences in each individual genome consist of a set variable genes. While some of the genetic variation can be explained with divergence caused by geographical isolation the high number of non-orthologous genes suggests the existence of a symbiont population that is exchanging genes with others. This circumstance is demonstrated by the difference in genome size of over 500,000 bp between the smallest and the largest of the genomes. Free-living populations have been found to possess far more variable genomes than genetically isolated intracellular symbionts [18][101].

Some genetic features found in the genome annotation do not fit a strictly intracellular lifestyle. All symbiont genomes have a complete gene set for the synthesis and utilization of a flagellum. In addition, several genes connected with bacterial chemotaxis as well as aerotaxis sensor receptor proteins were found in all genomes. Such capabilities for mobility are not expected in an obligate endosymbiont as there would be no need for motility within a bacteriocyte and the large number of unused genes would be an energetic burden for the cell. In the open marine environment however, chemotaxis can provide a number of advantages. Besides the ability to actively search for nutrients, flagella have been reported to facilitate the infection of hosts as is the case for the bacterium *Legionella pneumophila* the pathogenic agent for the Legionnaires' disease [103]. Chemotaxis may also be an enabling factor for symbiont-host recognition, however no specific responsible signaling molecules have been identified yet. Type IV pili, such as those conserved amongst all

symbiont genomes, have been shown to play a role in the attachment, surface movement and invasion of host tissue [104]. These type IV pili have also been connected with biofilm formation [105], however no genes for the excretion of extracellular matrix proteins or polysaccharides were found in the genomes.

These and other genomic features may have lead the PICA phylotyping software to classify the symbiont as "non-obligate intracellular". However, the authors themselves stated that the discrimination between obligate and facultative intracellular species is not optimal as the transition between the two lifestyles is fluent [61].

The host's reliance on the symbionts nutrient contribution naturally puts a strong selective pressure on the maintenance of the relationship. A lateral acquisition of symbionts decouples the genomic evolution of host and symbiont. As environmental bacteria, the symbionts would have more opportunities for genetic exchange through horizontal gene transfer with other microorganisms than any strictly intracellular microorganisms [16][18]. In some cases, acquired genes may provide additional benefits for the fitness of both the symbiont and the host.

9.5 Analysis of the marine sediment

The analysis of the genome revealed many clues about a symbiont habitat outside the clam (see chapter 9.4 above). Therefore, the next goal was to detect signs of the symbiont in environmental samples. The most likely candidates as habitat would be the marine sediment and the pore water where the clam host can be found. Sediment cores were taken from depths of up to 60 cm, just below the depths where we found the *Loripes* clams.

The sequencing of the shallow water marine sediment communities near the *Loripes orbiculatus* spot did not uncover any ribosomal DNA sequences related to *Cand.* Thiodiazotropha endoloripes. One explanation as to why the symbiont could not be found is a low abundance of these bacteria in the sediment samples. The chosen methods of Illumina amplicon deep-sequencing is biased by the common DNA extraction- and PCR-biases during sample preparation. DNA sequences not amplified in the preparatory PCR steps would not be present in the sequencing dataset. Additionally, with such approaches it is not

uncommon to discard up to half of all sequence reads during bioinformatical processing [106], which can lead to the exclusion of some rare taxa in the sample. Unfortunately, we cannot estimate the abundance and distribution of the symbiont in the environment using this method with the samples analyzed.

The analysis of the microbial community however, did uncover some reads related to the 16S rRNA gene of *Cand.* Thiodiazotropha endoloripes. The closest related sequences showed roughly 95% sequence similarity to the symbiont 16S sequence but could not be classified by the mothur software. One sequence of *Sedimenticola*, a close relative to the symbiont, was found in low numbers in the upper layers of the sediment core. This sequence is 93% similar to the *Loripes orbiculatus* and 95% similar to *Anodontia phillipiana* and *Solemya terraeregina* gill symbionts belonging to the same clade of unclassified sulfur-oxidizing gammaproteobacteria [13]. These sequences however are only between 425 - 450bp long, so their informative value is limited. The sediment in general contained many bacteria with the putative capacity to couple sulfur-oxidation with the ability of autotrophic growth, such as *Acidiferrobacter* [107], *Thiohalophilus* [108] and *Thiobacillus* [109]. Many of these autotrophs are also able to use nitrogen compounds as electron acceptors for respiration. Specifically, the globally spread genus *Acidiferrobacter* has been identified as one of the most important light-independent, carbon-fixing microorganisms in marine sediments. This clade contributes a large part of the carbon fixation in sediments, despite representing only a relatively small fraction of the total microbial community [110]. A shift in the community can be observed along the depth gradient. As a trend, it seems that sulfur-oxidizers and other oxygen dependent groups decrease in abundance deeper in the sediment. This may be due to a decrease in oxygen in the sediment’s fine graining, but oxygen saturation was not measured during the course of this study. It has to be mentioned here that the metabolic functions ascribed to members of the microbial clades are merely inferred by their 16S rRNA phylogeny. The attempt of linking the community composition to metabolic processes going on in the environment may be prone to errors. While this method does offer some predictive power, and may provide a basis for hypotheses it cannot replace classical experimental methods of assessing functions in an ecosystem like measurements of respiration or carbon and nitrogen fixation rates [111]. Another source of errors may be the wrong estimation of microbial community compositions based on ribosomal gene content in

the sample. An analysis of a microbial community can be biased by uneven extraction and PCR amplification rates, specifically for ribosomal RNA genes that can occur in varying copy numbers in a genome [112].

9.6 Secondary symbionts in *Loripes orbiculatus*

Binning and analysis of the *Loripes orbiculatus* metagenomes found 16S rRNA sequences of *Endozoicomonas* in the Phy3 metagenome and *Shewanella* sequences in the PhyM metagenomes, representing roughly 50% of all bacterial 16S rRNA reads in those two metagenomes respectively. Although this number could be inflated due to a possible increase in ribosomal gene content, it was still a surprising result. As these organisms were not a focus of this study, their possible role in the clam, based on the annotated genome, is only briefly discussed.

9.6.1 *Endozoicomonas*

Bacteria belonging to the genus *Endozoicomonas* have been detected as endosymbionts with largely unknown functions within a wide variety of marine animal species, ranging from simple invertebrates, to highly developed vertebrates such as fish [113]. Similar to its host range, this ubiquitously spread group also exhibits very high genome plasticity and large genome size [114], indicating a putative free-living stage and is suspected to harbor both beneficial symbionts as well as opportunistic parasites and pathogens [113].
The *Endozoicomonas* bin in this study originates from one clam individual from Mauritania and could not be assigned to a specific species through a BlastN search. Interestingly, bacteria of this group were only found within the Phy3 metagenome, although both Phy1 and Phy2 clams were collected from the same sampling site in Piran. It is possible that this particular strain of *Endozoicomonas* may specifically infect only a certain strain of hosts.
The annotation of the genome revealed some of the genetic potential of this organism giving clues to its possible interactions with the host. *Endozoicomonas* does not seem to encode genes for an ubiquinol oxidase required for aerobic respiration, but is capable of

fermentation and anaerobic respiration. This may prevent competition for oxygen between the primary symbiont and *Endozoicomonas*. A large part of its metabolism seems to be dedicated to siphoning nutrients such as amino acids, peptides and many organic compounds from the host. It remains unclear whether *Endozoicomonas* provides any metabolites in return but the capacity to do so may be provided in the form of secretion systems I, II and III. However, it is clear that this *Endozoicomonas* strain does not possess any metabolic capabilities for the fixation of inorganic carbon. One adaption to the heterotrophic lifestyle is the glyoxylate cycle, a modified version of the tricarboxylic acid(TCA) cycle [115]. This metabolic pathway allows the bacterium to use less complex carbon sources in a nutrient deprived environment and may be an adaption to survive phagocytosis through a host's immune system [116]. In addition to a type IV pilus, several genes for aggregation and biofilm formation may play a role in infection. Although the bacteria have not been observed under the microscope the presence of *RodA* as well as *MreBCD* genes suggests a rod-shaped cell habitus.

9.6.2 *Shewanella*

Similar to *Endozoicomonas*, *Shewanella* species have been found as facultative symbionts in a large number of marine eukaryotes. Members of this genus, in particular *S.oneidensis* have been studied for their extensive capabilities of reducing several oxidized metals, as well as sulfur compounds for respiration [117].

Much like *S. oneidensis*, the *Shewanella* strain in this study is facultative aerobic. It possesses the full gene set for the synthesis and utilization of a flagellum as well as several chemotaxis-related genes indicating an extracellular stage in its lifecycle. Similar to *Endozoicomonas*, *Shewanella* has no carbon-fixation capabilities and relies on a heterotrophic lifestyle being able to use the glyoxylate cycle to adapt to nutrient poor environments. Its ability to exploit a host for its organic nutrients however, is less pronounced. One reason why the lucinid gills might be a favorable environment for *Shewanella* could be its putative ability to use stored sulfides for respiratory processes. Its genome contains a gene cluster encoding the three subunits of a membrane bound polysulfide reductase. This enzyme, capable of reducing polysulfides to produce hydrogen sulfide [118], is required for sulfur as well as thiosulfate

respiration [119] and may provide access to the sulfide stored in globules within the lucinid gills.

Thus, *Shewanella* is well equipped to thrive in the sulfur-rich, microaerobic environment of the lucinid gills. It remains unclear whether *Shewanella* also offers any benefits for its host in exchange or acts as a pathogen, merely siphoning off metabolites.

10. Conclusions and outlook

10.1 Metabolism

Despite a research history of four decades, chemoautotrophic microorganisms, especially those who have entered a partnership with eukaryotic hosts, continue to demonstrate unexpected new traits. In this study, the genetic variability of a lucinid endosymbiont species was demonstrated and their metabolic capacities were explored. While some of the more important functions are already known from previous studies, the examination of genomes continues to turn up novel features. Many of the newly found genes are still missing functional annotations. Some of these features may have large impacts on the geochemistry in the bacterium's immediate environment and the biochemistry of its host. In this study, the genes with known functions alone have already shown that the different symbiont strains may have adaptions to different environments. This became very apparent when looking at the large variability in the symbionts nitrogen metabolism. This study, however, only shows the genomic potential of these organisms. An experimental confirmation of actual gene expression and utilization is still required before any conclusions about the symbionts ecophysiology can be drawn. It is a well-known fact that organisms do not necessarily use all the genes encoded in their genome. This is even more noticeable in microorganisms, as they are able to exchange parts of their genome with other organisms. Due to the inability to cultivate the symbiont without a host, cultivation-independent methods, such as enzyme assays, the analysis of stable isotope signatures or mass-spectrometry based approaches, like NanoSIMS, will be required to close knowledge gaps in this regard. Further, the study of gene expression trough metatranscriptomics and metaproteomics under varying conditions, such as anoxia or increased sulfur or nitrogen saturation, will be able to reveal the system's response to environmental changes.

10.2 Finding the symbiont in the environment

One of the major goals of this study was the search for evidence of a free-living stage of the lucinid symbiont. Unfortunately, this goal has not been met satisfactorily during this study. The detection of labeled cells under the fluorescence microscope does not provide absolute proof of the bacterium being environmental, however, this study did provide convincing arguments that the symbiont also exists outside the clam.

The insufficient specificity of the oligonucleotide probe 845-Cy3 created uncertainty about the correctness of the detected FISH signal in Figure 6. Therefore, I recommend the design of a new FISH-oligonucleotide probe that does not target any bacterial outside groups from the marine environment, before continuing the search for the symbiont. The Probe Design tool of the ARB software package [120] may be a useful resource for this purpose.

In the genome analysis, it became clear that the symbiont lacks genes for the synthesis and secretion of extracellular polysaccharides and matrix proteins required for the formation of biofilms. This may impede the symbionts ability to attach to surfaces. Instead, the bacterium may be found swimming freely in the water column of the pore water rather than attached to the sediment. This warrants a further analysis of the pore water with FISH which was begun in this study. The number of pore water signals detected by the FISH analysis in this study was quite low (less than 1% of all cells), suggesting that, if it was indeed the symbiont, it may be part of the rare biosphere in that environment. To counteract this, I suggest filtering larger volumes of pore water for screening, increasing cell density. This would also require a more stringent pre-filtering of the water, as a larger volume naturally also contains more debris and contaminants, especially in the lower layers of the sediment. In this study, the pore water filters from below 20 cm sediment depth proved to be heavily contaminated with debris. In such instances techniques of signal amplification such as catalyzed reporter deposition (CARD)-FISH may help to provide the necessary contrast against the strong background auto-fluorescence. As the amplicon sequencing of the sediment community suggested that putative oxygen dependent microorganisms exhibit a decrease in abundance the lower the depth in the sediment, efforts should be focused on the pore water from upper layers.

It is also worth considering that the presence of the symbiont may be very tightly linked to the presence of the lucinids, causing it to be found only in close proximity of the lucinids themselves. It may be useful to analyze the sediment and pore water directly in contact with the clams. There were no lucinid individuals found in the sediment cores collected from Elba as they were taken from an undisturbed, even spot in the sediment. Therefore, an analysis of their immediate habitat was not possible.

10.3 Other aspects of the symbiosis

While this study covered several aspects of the *Loripes orbiculatus* symbiosis, many open questions about the characteristics of this association remain. For example, the evolutionary advantages for the endosymbiont in this relationship are still questionable. It seems obvious that the symbiosis is obligate for the lucinids, as a large supply of nutrients comes from their symbionts and they benefit from the detoxification of sulfur compounds. The advantages for the symbiont on the other hand are less pronounced. The protection and environmental conditions provided by the gills allow the symbionts to grow in large density, however it is still unclear whether the symbionts are ever able to escape their host and re-enter the outside environment as is the case with the chemoautotrophic endosymbiont of the tubeworm *Riftia pachyptila* [121]. The association with the host may prove to be a dead end for these bacteria from an evolutionary perspective. If this were the case, we would expect the symbiont to develop defense mechanisms to avoid exploitation, however there are no signs of any countermeasures and the association seems evolutionary stable.

The co-evolution of lucinids and their symbionts is a wide field that offers many interesting research questions. Studies conducted by Gros et al. [78] as well as this study have shown that the symbiont is not restricted to one host and not as host-specific as previously thought. Instead, the present study found indications that symbiont strains are more specific to a certain habitat than to a host species. Similar results were delivered by a study from Brissac et al. from 2016 [122]. In this study, the symbiont strain diversities in lucinids from different habitats were compared. It was revealed that the same species of lucinid hosts harbored different strains of symbionts in different locations.

On the other hand, we observe a low strain heterogeneity when looking at the symbiont diversity within a single host individual. The gills of lucinids seem to be largely dominated by only one 16S rRNA phylotype of symbionts. This indicates a yet unknown selection process for the colonization of the gills. It is not clear whether this stems from the lucinids preference for only one phylotype of symbiont or competitive selection within the symbiont population, offering another interesting question to be answered.

11. References

1. Banse K. 1964. “On the vertical distribution of zooplankton in the sea”. *Prog. Oceanog.* 2: 53 – 125.

2. Cavanaugh C.M. et al. 2006. “Marine Chemosynthetic Symbioses”. In *The Prokaryotes: A Handbook on the Biology of Bacteria* (eds Dworkin M et al.) Springer, New York: 475 – 507.

3. Dublier N., Bergin C., Lott C. 2008. “Symbiotic diversity in marine animals: the art of harnessing chemosynthesis”. *Nat. Rev. Microbiol.* 6: 725 – 740.

4. Cavanaugh C.M. et al. 1981. “Prokaroytic cells in the hydrothermal vent tube worm Riftia pachyptila Jones: Possible chemoautotrophic symbionts”. *Science* 213:340 – 342.

5. Petersen J.M. et al. 2016. “Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation”. *Nat. microbiol*. 2: 16195.

6. Markert S. et al. 2011. “Status quo in physiological proteomics of the uncultured Riftia pachyptila endosymbiont”. *Proteomics* 11(15): 3106 – 3117.

7. Friedrich C.G. et al. 2005. “Prokaryotic Sulfur Oxidation”. *Curr. Opin. Microbiol.* 8: 253 – 259.

8. Felbeck H. and Jarchow J. 1998. “Carbon Release from Purified Chemoautotrophic Bacterial Symbionts of the Hydrothermal Vent Tubeworm Riftia pachyptila”. *Physiol. Zool.* 71(3): 294 – 302.

9. Bright M. and Lallier F.H. 2010. “The biology of vestimentiferan tubeworms”. In *Oceanography and Marine Biology: An Annual Review* 48: 213 – 265. Gibson RN, Atkinson RJA and Gordon JDM, eds. CRC Press.

10. Bright M., Keckeis H. and Fisher C.R. 2000. “An autoradiographic examination of carbon fixation, transfer and utilization in the Riftia pachyptila symbiosis”. *Mar. Biol.* 136: 621 – 632.

11. Hentschel U. and Felbeck H. 1993. “Nitrate respiration in the hydrothermal vent tubeworm Riftia pachyptila”. *Nature* 366: 338 – 340.

12. Hentschel U., Cary S.C. and Felbeck H. 1993. “Nitrate respiration in chemoautotrophic symbionts of the bivalve Lucinoma aequizonata”. *Mar. Ecol. Progr. Ser.* 94: 35 – 41.

13. Cavanaugh C.M. 1985. “Symbioses of chemoautotrophic bacteria and marine invertebrates from hydrothermal vents and reducing sediments”. *Biol. Soc. Wash. Bull.* 6: 373 – 388.

14. Conway N.M. and Capuzzo J.E.M. 1992. “High taurine levels in the Solemya velum symbiosis”. *Comp. Biochem. Physiol. B Biochem. Melec. Biol.* 102: 175 – 185.

15. Childress J.J. et al. 1991. “Sulfide-driven autotrophic balance in the bacterial symbiont-containing hydrothermal vent tubeworm Riftia pachyptila Jones”. *Biol. Bull.* 180: 135 – 153.

16. McCutcheon J.P. and Moran N.A. 2012. “Extreme genome reduction in symbiotic bacteria”. *Nat. Rev.* 10: 13 – 26.

17. Peek A.S., Vrijenhoek R.C. and Gaut B.S. 1998. “Accelerated Evolutionary Rate in Sulfur-Oxidizing Endosymbiotic Bacteria Associated with the Mode of Symbiont Transmission”. *Mol. Biol. Evol.* 15(11): 1514 – 1523.

18. Moran N.A. 1996. “Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria”. *Proc. Natl. Acad. Sci. USA.* 93: 2873 – 2878.

19. Moran N.A. and Bennett G.M. 2014. “The Tiniest Tiny Genomes”. *Annu. Rev. Microbiol.* 68: 195 – 215.

20. Bennett G.M. and Moran N.A. 2013. “Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect”. *Genome Biol. Evol.* 5: 1675 – 1688.

21. Vogan A.A. and Higgs P.G. 2011. “The advantages and disadvantages of horizontal gene transfer and the emergence of the first species”. *Biology Direct* 6:1.

22. Bright M. and Bulgheresi S. 2010. “A complex journey: transmission of microbial symbionts”. *Nat. Rev. Microbiol.* 8(3): 218 – 230.

23. Dublier N. et al. 1995. “Characterization of Chemoautotrophic Bacterial Symbionts in a Gutless Marine Worm (Oligochaeta, Annelida) by Phylogenetic 16S rRNA Sequence Analysis and In Situ Hybridization”. *Appl. Environ. Microbiol.* 61(6): 2346 – 2350.

24. Distel D.L. et al. 2017. “Discovery of chemoautotrophic symbiosis in the giant shipworm Kuphus polythalmia (Bivalvia: Teredinidae) extends wooden-steps theory”. *PNAS* 114(18): E3652–E3658, doi: 10.1073/pnas.1620470114

25. Woyke T. et al. 2006. “Symbiosis insights through metagenomic analysis of a microbial consortium”. *Nature* 443: 950 – 955.

26. Caro A. et al. 2007. “Characterization of the Population of the Sulfur-Oxidizing Symbiont of Codakia orbicularis (Bivalvia, Lucinidae) by Single-Cell Analyses”. *Appl. Environ. Microbiol.* 2101 – 2109.

27. Gros O., Liberge M. and Felbeck H. 2003. “Interspecific infection of aposymbiotic juveniles of Codakia orbicularis by various tropical lucinid gill-endosymbionts”. *Mar. Biol.* 142: 57 – 66.

28. Taylor J.D. and Glover E.A. 2006. “Lucinidae (Bivalvia) – the most diverse group of chemosymbiotic mollusks”. *Zool. J. Linnean Soc.* 148: 421 – 438.

29. Frenkiel L. and Mouëza M. 1995. “Gill ultrastructure and symbiotic bacteria in Codakia orbicularis (Bivalvia, Lucinidae)”. *Zoomorphology* 115: 51 – 61.

30. Allen J.A. 1958. “On the basic form and adaptations to habitat in the Lucinacea (Eulamellibanchia)”. *Phil. Trans. R. Soc. B* 241: 421 – 484.

31. Herry A., Diouris M. and Le Pennec M. 1989. “Chemoautotrophic symbionts and translocation of fixed carbon from bacteria to host tissues in the littoral bivalve Loripes lucinalis (Lucinidae)”. *Mar.Biol.* 101: 305 – 312.

32. Distel D.L. and Felbeck H. 1987. “Endosymbiosis in the lucinid clams Lucinima aequizonata, Lucinoma annulata and Lucina floridana: a reexamination of the functional morphology of the gills as bacteria – bearing organs”. *Mar. Biol.* 96: 79 – 86.

33. Le Pennec M. et al. 1994. “Nutrition – gametogenesis relationship in the endosymbiont host bivalve Loripes lucinalis (Lucinidae) from reducing coastal habitats”. *Proc. 28th European Mar. Biol. Symp. Heraklion, Crete* (in press).

34. Johnson M.A. and Fernandez C. 2001. “Bacterial symbiosis in Loripes lucinalis (Mollusca: Bivalvia) with comments on reproductive strategy”. *J. Mar. Biol. Ass. U.K.* 81: 251 – 257.

35. Stanley S.M. 2014. “Evolutionary radiation of shallow-water Lucinidae (Bivalvia with endosymbionts) as a result of the rise of seagrasses and mangroves”. *Geology* 42: 803 – 806.

36. Espinosa E.P. et al. 2013. “Endosymbiotic bacteria in the bivalve Loripes lacteus: Localization, characterization and aspects of symbiont regulation”. *J. Exp. Mar. Biol. Ecol.* 448: 327 – 336.

37. König S. et al. 2016. “Nitrogen fixation in a chemoautotrophic lucinid symbiosis”. *Nat. Microbiol.* 2, 16193

38. Van der Heide T. et al. 2012. “A Three-Stage Symbiosis Forms the Foundation of Seagrass Ecosystems”. *Science* 336: 1432 – 1434.

39. Reynolds L.K., Berg P. and Zielman J.C. 2007. “Lucinid Clam Influence on the Biogeochemistry of the Seagrass Thalassia testudinum Sediments”. *Estuaries Coasts* 30: 482.

40. Higgs N.D., Newton J. and Attrill M.J. 2016. “Caribbean Spiny Lobster Fishery Is Underpinned by Trophic Subsidies from Chemosynthetic Primary Production”. *Curr. Biol.* 26(24): 3393 – 3398.

41. Steedman H.F. 1975. “Polyester Wax: A New Ribboning Embedding Medium for Histology”. *Nature* 179: 1345.

42. Müller M.R.A., Ehrmann M.A. and Vogel R.F. 2000. „Multiplex PCR for the Detection of Lactobacillus pontis and Two Related Species in a Sourdough Fermentation”. *Appl. Environ. Microbiol.* 66(5) 2113 – 2116.

43. Brandl M.T., Quinones B. and Lindow S.E. 2001. “Heterogenous transcription of an indoleacetic acid biosynthetic gene in Erwinia herbicola on plant surfaces”. *PNAS.* 98 (6): 3454 – 3459.

44. Brown P.B. and Wolfe G.V. 2006. “Protist Genetic Diversity in the Acidic Hydrothermal Environments of Lassen Volcanic National Park, USA”. *J. Eukaryot. Microbiol.* 53: 420 – 431.

45. Medlin L. et al. 1988. “The characterization of enzymatically amplified eukaryotic 16S-like rRNA coding regions”. *Gene* 71: 491 – 499.

46. Herlemann D.P.R. et al. 2011. “Transitions in the bacterial communities along the 2000 km salinity gradient of the Baltic Sea”. *The ISME Journal* 5(10): 1571 – 1579.

47. Manz W. et al. 1992. “Phylogenetic Oligodeoxynucleotide Probes for the Major Subclasses of Proteobacteria: Problems and Solutions”. *System. Appl. Microbiol.* 15: 593 – 600.

48. Wallner G., Amann R. and Beisker W. 1993. “Optimizing Fluorescent In Situ Hybridization With rRNA-Targeted Oligonucleotide Probes for Flow Cytometric Identification of Microorganisms”. *Cytometry* 14: 136 – 143.

49. Geospiza. 2009. *FinchTV 1.4.0.* Geospiza, Inc., Seattle, WA. http://www.geospiza.com.

50. Altschul S.F. et al. 1990. „Basic Local Alignment Search Tool”. *J. Mol. Biol.* 215(3): 403 – 410.

51. Tamura K. et al. 2013. “MEGA6 : Molecular Evolution Genetics Analysis Version 6.0”. *Mol. Biol. Evol.* 30: 2725 – 2729.

52. Gruber-Vodicka H., Pruesse E.A. and Seah B. https://github.com/HRGV/phyloFlash

53. Bushnell B. BBMap: Short read aligner. https://sourceforge.net/projects/bbmap/

54. Bankevich A. et al. 2012. “SPAdes: A New Genome Assembly Algorithm and its Application to Single-Cell Sequencing”. *J Comput. Biol.* 19: 455 – 477.

55. Seah B.K.B. and Gruber-Vodicka H.B. 2015. “Gbtools: Interactive Visualization of Metagenome Bins in R”. *Front. Microbiol.* 6: 1451.

56. Parks D.H. et al. 2015. “CheckM: assessing the quality of microbial genomes recovered from isaolates, single cells, and metagenomes”. *Genome Res.* 25: 1043 – 1055.

57. Aziz R.K. et al. 2008. “The RAST Server: Rapid Annotation using Subsystem Technology”. *BMC Genomics*. 9: 75.

58. Li H. et al. 2009. “The sequence alignment/map format and SAMtools”. *Bioinformatics* 25: 2078 – 2079.

59. Robinson J.T. et all. 2011. “Integrative Genomics Viewer”. *Nat. Biotech.* 29: 24 – 26.

60. MacDonald N.J. and Beiko R.G. 2010. “Efficient learning of microbial genotype-phenotype association rules”. *Bioinformatics*: 26(15): 1834 – 1840.

61. Feldbauer R. et al. 2015. “Prediction of microbial phenotypes based on comparative genomics”. *BMC Bioinformatics:* 16 Suppl. 14 S1.

62. Contreras-Moreira B. and Vinueas P. 2013. “GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis”. *Appl. Environ. Microbiol.* 79(24): 7696 – 7701.

63. Li L., Stoeckert C.J.Jr. and Roos D.S. 2003. “OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes”. *Genome Res.* 13: 2178 – 2189.

64. Lukjancenko O., Wassenaar T.M. and Ussery D.W. 2010. “Comparison of 61 Sequenced Escherichia coli Genomes”. *Microb. Ecol.* 60(4): 708 – 720.

65. Jacobsen A. et al. 2011. “The Salmonella enterica Pan-genome”. *Microb. Ecol.* 62: 487

66. Karp P.D., Paley S. and Romero P. 2002. “The pathway tools software”. *Bioinformatics* 18 (suppl. 1) : S225 – S232.

67. Herbold C.W. et. al. 2015. "A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes". *Front. Microbiol.*(6); 731.

68. Schloss P.D. et al. 2009. „Introducing mothur: Open – Source. Platform – Independent. Community – Supported Software for Describing and Comparing Microbial Communities". *Appl. Environ. Microbiol.* 75 (23): 7537 – 7541.

69. R Development Core Team. 2008. "R: A language and environment for statistical computing". R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

70. Oksanen J. et al. 2013. "Vegan: Community Ecology Package". R package version 2.4-3. Available at https://cran.r-project.org/web/packages/vegan/index.html

71. McMurdie P.J., Holmes S. 2013. "phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data". *PLoS ONE*: 8(4): e61217. doi:10.1371/journal.pone.0061217

72. Wickham H. "ggplot2: An implementation of the Grammar of Graphics 2008". Available at: http://had.co.nz/ggplot2/

73. Kimura M. 1980. "A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences". *J. Mol. Evol.* 16: 111 – 120.

74. Kleiner et al. 2012. "Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use". *PNAS* 109(19): E1173 – E1182.

75. Valentini M. et al. 2011. "Identification of C(4)-dicarboxylate transport systems in Pseudomonas aeruginosa PAO1". *J. Bacteriol.* 193(17): 4307 – 4316.

76. Wagner M., Horn M. and Daims H. 2003. "Fluorescence in situ hybridization for the identification of prokaryotes". *Curr. Opin. Microbiol.* 6: 302 – 309.

77. Loy A, Arnold R et. al. 2008. "ProbeCheck – a central resource for evaluating oligonucleotide probe coverage and specificity". *Environ. Microbiol.* 10 (10) 2894 – 2898.

78. Gros O. et al. 2012. "Plasticity of symbiont acquisition throughout the life cycle of the shallow-water tropical lucinid Codakia orbiculata (Mollusca: Bivalvia)". *Environ. Microbiol.* 14: 1584 – 1595.

79. Stackebrandt E. and Goebel B.M. 1994. "Taxonomic note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology". *Int J. Syst. Bacteriol.* : 846 – 849.

80. Mandal S.D. et al. 2014. "Mitochondrial markers for the identification and phylogenetic studies in insects – a review". *DNA Barcodes* 2(1). (doi:10.2478/dna-2014-0001)

81. Hebert P.D.N., Ratnasingham S. and deWaard J.R. 2003. "Barcoding animal life: cytochrome c subunit 1 divergences among closely related species". *Proc. Biol. Sci.* 270: 96 – 99.

82. Layton K.K.S., Martel A.L. and Hebert P.D.N. 2014. "Patterns of DNA Barcode Variation in Canadian Marine Molluscs". *PLoSmONE* 9(4): e95003. https://doi.org/10.1371/journal.pone.0095003

83. Jaspers E. and Overmann J. 2004. "Ecological Significance of Microdiversity: Identical 16S rRNA Gene Sequences Can Be Found in Bacteria with Highly Divergent Genomes and Ecophysiologies". *Appl. Environ. Microbiol.* 70 (8): 4831 – 4839.

84. Robidart J.C. et al. 2008. "Metabolic versatility of the Riftia pachyptila endosymbiont revealed through metagenomics". *Environ. Microbiol.* 10(3): 727 – 737.

85. Newton I.L. et al. 2007. "The Calyptogena magnifica chemoautotrophic symbiont genome". *Science* 315(5814): 998 – 1000.

86. Childress J.J. and Girguis P.R. 2011. "The metabolic demands of endosymbiotic chemoautotrophic metabolism on host physiological capabilities". *J. Exp. Biol.* 214: 312 – 325.

87. Girguis P.R. et al. 2000. "Fate of Nitrate Acquired by the Tubeworm Riftia pachyptila". *Appl. Environ. Ecol.* 66(7): 2783 – 2790.

88. An S. and Gardner W.S. 2002. "Dissimilatory nitrate reduction to ammonium (DNRA) as a nitrogen link, versus denitrification as a sink in a shallow estuary (Laguna Madre/Baffin Bay, Texas)". *Mar. Ecol. Prog. Ser.* 237: 41 – 50.

89. Girguis P.R. and Childress J.J. 2006. "Metabolite uptake, stoichiometry and chemoautotrophic function of the hydrothermal vent tubeworm Riftia pachyptila: responses to environmental variations in substrate concentrations and temperature". *J. Exp. Biol.* 209: 3516 – 3528.

90. Wagner M. et al. 2015. "Cyanate as energy source for nitrifiers". *Nature* 524(7563): 105 – 108.

91. Anderson P.M., Sung Y. and Fuchs J. A. 1990. "The cyanase operon and cyanate metabolism". *FEMS Microbiol. Rev.* 7(3 – 4): 247 – 252.

92. Duplessis M.R. et al. 2004. "Respiration Strategies utilized by the Gill Endosymbiont from the Host Lucinid Codakia orbicularis (Bivalvia: Lucinidae)". *Appl. Environ. Microbiol.* 70(7): 4144 – 4150.

93. Muth W.L. and Costilow R.N. 1974. "Ornithine Cyclase (Deaminating). II. Properties of the homogeneous enzyme". *J. Biol. Chem.* 249(23): 7457 – 7462.

94. Brandriss M.C. and Magasanik B. 1980. "Proline: an Essential Intermediate in Arginine Degradation in Saccharomyces cerevisiae". *J. Bacteriol.* 143(3): 1403 – 1410.

95. Murrell J.C. et al. 2015. "XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments". *Environ. Microbiol.* 17: 3937 – 3948.

96. Anthony C. 1982. "The Biochemistry of Methylotrophs". *Academic Press INC*, London.

97. Zhuang G-C. et al. 2014. "Gas chromatographic analysis of methanol and ethanol in marine sediment pore waters: Validation and implementation of three pretreatment techniques. *Mar. Chem.* 160: 82 – 90.

98. Hanson R.S. and Hanson T.E. 1996. "Methanotrophic bacteria". *Microbiol Rev.* 60(2): 439 – 471.

99. Payne W.J. 1973. "Reduction of nitrous oxides by microorganisms". *Bacteriol. Rev.* 37(4): 409 – 452.

100. Koonin E.V. and Wolf Y.I. 2008. "Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world". *Nucleic Acids Res.* 36(21): 6688 – 6719.

101. Ochman H. and Davalos L.M. 2006. "The Nature and dynamics of Bacterial genomes". *Science* 311(5768): 1730 – 1733.

102. Taylor J.D. and Glover E. 2000. "Functional anatomy, chemosymbiosis and evolution of the Lucinidae". In Harper E. M., Taylor J. D. and Crame J. A. "The Evolutionary Biology of the Bivalvia: Geological Society of London Special Publication" 177: 207 – 225.

103. Dietrich C. et al. 2001. "Flagellum of Legionella pneumophila Positively Affects the Early Phase of Infection of Eukaryotic Host Cells". *Infect. Immun.* 69(4): 2116 – 2122.

104. Wenyuan S. and Hong S. 2002. "Type IV Pilus-Dependent Motility and Its Possible Role in Bacterial Pathogenesis". *Infect Immun.* 70(1): 1 – 4.

105.O'Toole G.A. and Kolter R. 1998. "Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* Biofilm formation". 30(2): 295 – 304.

106. Eiler A. et al. 2015. "Microbial Community Composition and Diversity via 16S rRNA Gene Amplicons: Evaluating the Illumina Platform". *PLoS ONE* 10(2): e0116955. doi: 10.1371/journal.pone.0116955.

107. Hallberg K.B., Hedrich S. and Johnson D.B. 2011. "Acidiferrobacter thiooxydans, gen. nov. sp. nov.; an acidophilic, thermo-tolerant, facultatively anaerobic iron- and sulfur-oxidizer of the family Ectothiorhodospiraceae". *Extremophiles* 15(2): 271 – 279.

108. Sorokin D.Y. et al. 2007. "Denitrification in a binary culture and thiocyanate metabolism in Thiohalophilus thiocyanoxidans gen. nov. sp. nov. – a moderately halophilic chemolithoautotrophic sulfur-oxidizing Gammaproteobacterium from hypersaline lakes". *Arch. Microbiol.* 187(6) : 441 – 450.

109. Robertson L.A. and Kuenen J.G. 2006. "The Genus Thiobacillus". In "*The Prokaryotes"*, part 1, section 3.2: 812 – 827.

110. Mußmann et al. 2016. "Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments". *ISME J.* 10(8): 1939 – 1953.

111. Cavender-Bares J. et al. 2009. "The merging of community ecology and phylogenetic biology". *Ecol. Lett.* 12: 693 – 715.

112. Stackebrandt E., Rainey F.A. and Farrelly V. 1995*ppl. Environ. Microbiol.* 61(7): 2798 – 2801.

113. Neave M.J. et al. 2016. "Diversity and function of prevalent symbiotic marine bacteria in the genus Endozoicomonas". *Appl. Microbiol. Biotechnol.* 100: 8315 – 8324.

114. Neave M.J. et al. 2017. "Endozoicomonas genomes reveal functional adaption and plasticity in bacterial strains symbiotically associated with diverse marine hosts". *Sci. Rep.* 7, 40579; doi: 10.1038/srep40579(2017).

115. Kondrashov F.A. et al. 2006. "Evolution of glyoxylate cacla enzymes in Metazoa: evidence of multiple horizontal transfer events and pseudogene formation". *Biology Direct* 1: 31. doi:10.1186/1745-6150-1-31

116. Lorenz M.C. and Fink G.R. 2002. "Life and Death in a Macrophage: Role of the Glyoxylate Cycle in Virulence". *Eukaryot. Cell* 1(5): 657 – 662.

117. Heidelberg J.F. et al. 2002. "Genome sequence of the dissimilatory metal ion – reducing bacterium Shewanella oneidensis". *Nature Biotechnol.* 20: 1118 – 1123.

118. Jormakka M. et al. 2008. “Molecular mechanism of energy conservation in polysulphide respiration”. *Nat. Struct. Mol. Biol.* 15(7): 730 – 737.

119. Burns J.L. and DiChristina T.J. 2009. “Anaerobic Respiration of Elemental Sulfur and Thiosulfate by Shewanella oneidensis MR-1 Requires psrA, a Homolog of the phsA Gene of Salmonella enterica Serovar Typhimurium LT2”. *Appl. Environ. Microbiol.* 75(16): 5209 – 5217.

120. Ludwig W. et al. 2004. “ARB: a software environment for sequence data”. *Nucleic Acids Res.* 32(4): 1363 – 1371.

121. Klose J. et al. 2015. “Endosymbionts escape dead hydrothermal vent tubeworms to enrich the free living population”. *PNAS* 112(36): 11300 – 11305.

122. Brissac T. et al. 2016. “Unexpected structured intraspecific diversity of thioautotrophic bacterial gill endosymbionts within the Lucinidae (Mollusca: Bivalvia).” *Mar. Biol.* 163(8): 176.

123. Probe designation according to Alm E W. et al. 1996. “The oligonucleotide probe database.” *Appl. Environ. Microbiol.* 62(10): 3557 – 3559.

12. Appendix

12.1 Supplementary Materials and Methods

12.1.1 DNeasy Blood & Tissue kit protocol (QIAGEN)

Protocol: Purification of total DNA from Animal Tissues (Spin-Column Protocol)

1. Cut up to 25 mg into small pieces and place in a 1,5 ml microcentrifuge tube. Add 180 µl Buffer ATL.
2. Add 20 µl proteinase K. Mix thouroughly by vortexing, and incubate at 56°C until the tissue is completely lysed. Gill tissue samples were lysed over-night. Vortex occasionally during incubation to disperse the sample, or place in a thermomixer, shaking water bath, or on a rocking platform.
3. Vortex for 15 s. Add 200 µl Buffer AL to the sample, and mix thouroughly by vortexing. Then add 200 µl ethanol (96-100%), and mix again thoroughly by vortexing.
4. Pipet the mixture from step 3 (including any precipitate) into DNeasy Mini spin column placed in a 2 ml collection tube (provided). Centrifuge at ≥6000 x g (8000 rpm) for 1 min. Discard flow-through and collection tube.
5. Place DNeasy Mini spin column in a new 2 ml collection tube (provided), add 500 µl Buffer AW1, and centrifuge for 1 min at ≥6000 x g (8000 rpm). Discard flow through and collection tube.
6. Place DNeasy Mini spin column in a new 2 ml collection tube (provided), add 500 µl Buffer AW2, and centrifuge for 3 min at 20.000 x g (14.000 rpm) to dry the DNeasy membrane. Discard flow through and collection tube.
7. Place the DNeasy Mini spin column in a clean 1,5 ml or 2 ml microcentrifuge (not provided), and pipet 200 µl Buffer AE directly onto the DNeasy membrane. Incubate at room temperature for 1 min, and the centrifuge for 1 min at ≥6000 x g (8000 rpm) to elute
8. Recommended: For maximum DNA yield, repeat elution once as described in step 7.

12.1.2 Power Soil® DNA Isolation kit protocol (MO BIO Laboratories Inc.)

1. To the PowerBead Tubes provided, add 0.25 grams of soil sample.
2. Gently vortex to mix
3. Check Solutions C1. If Solution C1 is precipitated, heat solution to 60°C until dissolved before use
4. Add 60 µl of Solution C1 and invert several times or vortex briefly
5. Secure PowerBead Tubes horizontally using the MO BIO Vortex Adapter tube holder for the vortex (MO BIO Catalog# 13000-V1-24) or secure tubes horizontally on a flat bed vortex pad with tape. Vortex at maximum speed for 10 minutes. Note: If you are using the 24 place Vortex Adapter for more than 12 preps, increase the vortex time by 5-10 minutes.
6. Make sure the PowerBead Tubes rotate freely in your centrifuge without rubbing. Centrifuge tubes at 10,000 x g for 30 seconds at room temperature. CAUTION: Be sure not to exceed 10,000 x g or tubes may break.
7. Transfer the supernatant to a clean 2 ml Collection Tube (provided). Note: Expect between 400 to 500 µl of supernatant. Supernatant may still contain some soil particles.
8. Add 250 µl of Solution C2 and vortex for 5 seconds. Incubate at 4°C for 5 minutes.
9. Centrifuge the tubes at room temperature for 1 minute at 10,000 x g.
10. Avoiding the pellet, transfer up to, but no more than, 600 µl of supernatant to a clean 2 ml Collection Tube (provided).
11. Add 200 µl of Solution C3 and vortex briefly. Incubate at 4°C for 5 minutes.
12. Centrifuge the tubes at room temperature for 1 minute at 10,000 x g.
13. Avoiding the pellet, transfer up to, but no more than 750 µl of supernatant into a clean 2 ml Collection Tube (provided)
14. Shake to mix Solution C4 before use. Add 1200 µl of Solution C4 to the supernatant and vortex for 5 seconds
15. Load approximately 675 µl onto a Spin Filter and centrifuge at 10,000 x g for 1 minute at room temperature. Discard the flow through and add an additional 675 µl of supernatant to the Spin Filter and centrifuge at 10,000 x g for 1 minute at room temperature. Load the remaining supernatant onto the Spin Filter and centrifuge at 10,000 x g for 1 minute at room temperature.
16. Add 500 µ of Solution C5 and centrifuge at room temperature for 30 seconds at 10,000 x g
17. Discard the flow through
18. Centrifuge again at room temperature for 1 minute at 10,000 x g.
19. Carefully place spin filter in a clean 2 ml Collection Tube (provided). Avoid splashing any Solution C5 onto the Spin Filter.
20. Add 100 µl of Solution C6 to the center of the white filter membrane. Alternatively, sterile DNA-Free PCR Grade Water may be used for elution from the silica Spin Filter membrane at this step.
21. Centrifuge at room temperature for 30 seconds at 10,000 x g.
22. Discard the Spin Filter. The DNA in the tube is now ready for any downstream application. No further steps are required.

12.1.3 QIAquick PCR purification kit protocol (QIAGEN)

- Add ethanol (96 – 100%) to Buffer PE before use (see bottle label for volume).
- All centrifugation steps are carried out at 17,900 x g (13,000 rpm) in a conventional tabletop microcentrifuge at room temperature (15 – 25°C).

1. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix. It is not necessary to remove mineral oil or kerosene.
2. Place a QIAquick spin column in a provided 2 ml collection tube.
3. To bind DNA, apply the sample to the QIAquick column and centrifuge for 30 – 60 s.
4. Discard flow-through. Place the QIAquick column back into the same tube.
5. To wash, add 0,75 ml Buffer PE to the QIAquick column and centrifuge for 30 – 60 s.
6. Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column an additional 1 min.
7. Place QIAquick column in a clean 1,5 ml microcentrifuge tube.
8. To elute DNA, add 50 µl Buffer EB (10 mM Tris, pH 8,5) or water (pH 7,0 – 8,5) to the center of the QIAquick membrane and centrifuge the column for 1 min. Alternatively, for increased DNA concentration, add 30 µ elution buffer to the center of the QIAquick membrane, let the column stand for 1 min, and the centrifuge.

12.1.4 ZR-96 DNA Clean-up kit protocol (ZYMO Research Corp.)

1. In a 1,5 ml microcentrifuge tube, add 2 volumes of DNA Binding Buffer to each volume of DNA sample. (e.g., 300 µl binding buffer to 150 µl sample). Mix briefly by vortexing.
2. Transfer sample mixtures to the wells of a Silicon-A™ Plate mounted onto a Collection Plate.
3. Centrifuge at ≥ 3,000 x g (5,000 x g max.) for 5 minutes until sample mixtures have been completely filtered. Discard the flow-through.
4. Add 300 µl Wash Buffer to each well of the Silicon-A™ Plate. Centrifuge at ≥ 3,000 x g for 5 minutes. Repeat wash step.
5. Add 30 – 40 µl water directly to the column matrix in each well. Transfer the Silicon-A™ Plate onto an Elution Plate and centrifuge at ≥ 3,000 x g for 3 minutes to elute DNA. Ultra-pure DNA in water is now ready for use.

12.1.5 DNA quantification with Quant-iT™PicoGreen® dsDNA reagents kit (Thermo Fisher Scientific)

This protocol is based on the manufacturers manual. Standards and samples were measured in triplicates taking the mean values for concentration calculation. The assays were prepared in black 96-well-microplates (Greiner Bio-One, Kremsmünster, AUT), measurement were carried out with an Infinite M200 microplate reader (Tecan Group AG, Männerdorf CH) .

1. Prepare 1 x TE buffer by diluting 20 x TE buffer 1:20.
2. Dilute the Standard solution (λ DNA 100 ng/µl) to a concentration of 2 ng/ µl
3. Prepare PicoGreen® working solution: 9950 µl 1 x TE + 50 µl PicoGreen®
4. Prepare the following Standard mixture in the first 3 columns of the plate:

	Std. (2 ng)	**1 x TE buffer**	**Final amount (ng)**
A	50 µl	50 µl	100
B	37,5 µl	62,5 µl	75
C	25 µl	75 µl	50
D	12,5 µl	87,5 µl	25
E	5 µl	95 µl	10
F	2,5 µl	97,5 µl	5
G	1,2 µl	98,8 µl	2,5
H	0 µl	100 µl	Blank

5. Pipette 1 µl of each sample and 99 µl of 1 x TE buffer in triplicate in remaining wells.
6. Pipette 100 µl of PicoGreen® work solution in each well.
7. Protect plate from light and incubate for 2 – 3 min at room temperature.
8. Read plate according to the following conditions:

Excitation	~ 480 nm
Emission	~ 520 nm
Integration time	40 s
Lag time	0 s
Gain	Optimal
Number of flashes	10
Claculated well	Highest standard
Shaking	5 s

9. Subtract the fluorescence value of the reagent blank from that of each of the samples. Use corrected data to generate a standard curve of fluorescence versus DNA concentration.
10. Insert the mean of the sample fluorescent values into the function of the standard curve. The resulting value is the DNA concentration in ng/µl

12.1.6 SOP- Embedding in Steedman's wax for FISH according to Steedman (1957)

Preparation of wax:

1) Mix 1-hexadecanol with polyethylene glycol distearate 1:9 (e.g. 3 g 1-hexadecanol, 27 g PEG distearate in a 50 ml Falcon tube)
2) Melt in a 60°C water bath, mix well after wax has melted
3) Store in a water bath or oven at 38 – 40°C until needed. Wax can be left overnight at this temperature, if it is not needed for longer than this, it can be left to harden at room temperature and re-melted at 60°C before use

Embedding

1) Pre-warm wax and ethanol in the oven (38°C)
2) Dehydrate tissue :
    1. 70% ETOH for 30 min at RT
    2. 80% ETOH for 30 min at RT
    3. 96% ETOH for 30 min at RT
    4. 96% ETOH/Wax (3:1) for 60 min at 38°C — Steps 4 and 5 are optional
    5. 96% ETOH/Wax (2:1) for 60 min at 38°C
    6. 96% ETOH/Wax (1:1) for 60 min at 38°C
    7. Wax for 60 min at 38°C
    8. Wax for 60 min at 38°C
    9. Wax for 60 min at 38°C
3) Pre-warm heating plate at 35°C
4) Fill a beaker with 38°C water, put a thermometer and the embedding molds inside
5) Pour Steedman's wax into embedding molds, but don't fill completely
6) Place tissue in embedding mold, fill with wax
7) Let the block cool and harden over night at RT
8) Before cutting store the block at -20°C

12.1.7 Fluorescent *in situ* hybridization of gill tissue and filter sections

<u>Buffers and solutions</u>

5M NaCl
1M TRIS/HCl pH 8,0
0,5M EDTA pH 8,0
10% SDS sterile filtered

Washing buffer (for 50 ml):	1 ml 1M Tris/HCl 50 µl 10% SDS 0,5 ml 0,5M EDTA 0 - 45mM NaCl depending on formamide concentration Fill up to 50 ml with ddH_2O
Hybridization buffer (for 1 ml):	180 µl 5 M NaCl 20 µl Tris/HCl 1 µl 10% SDS 0 – 700 µl formamide Fill up to 1 ml with ddH_2O

<u>FISH procedure</u>

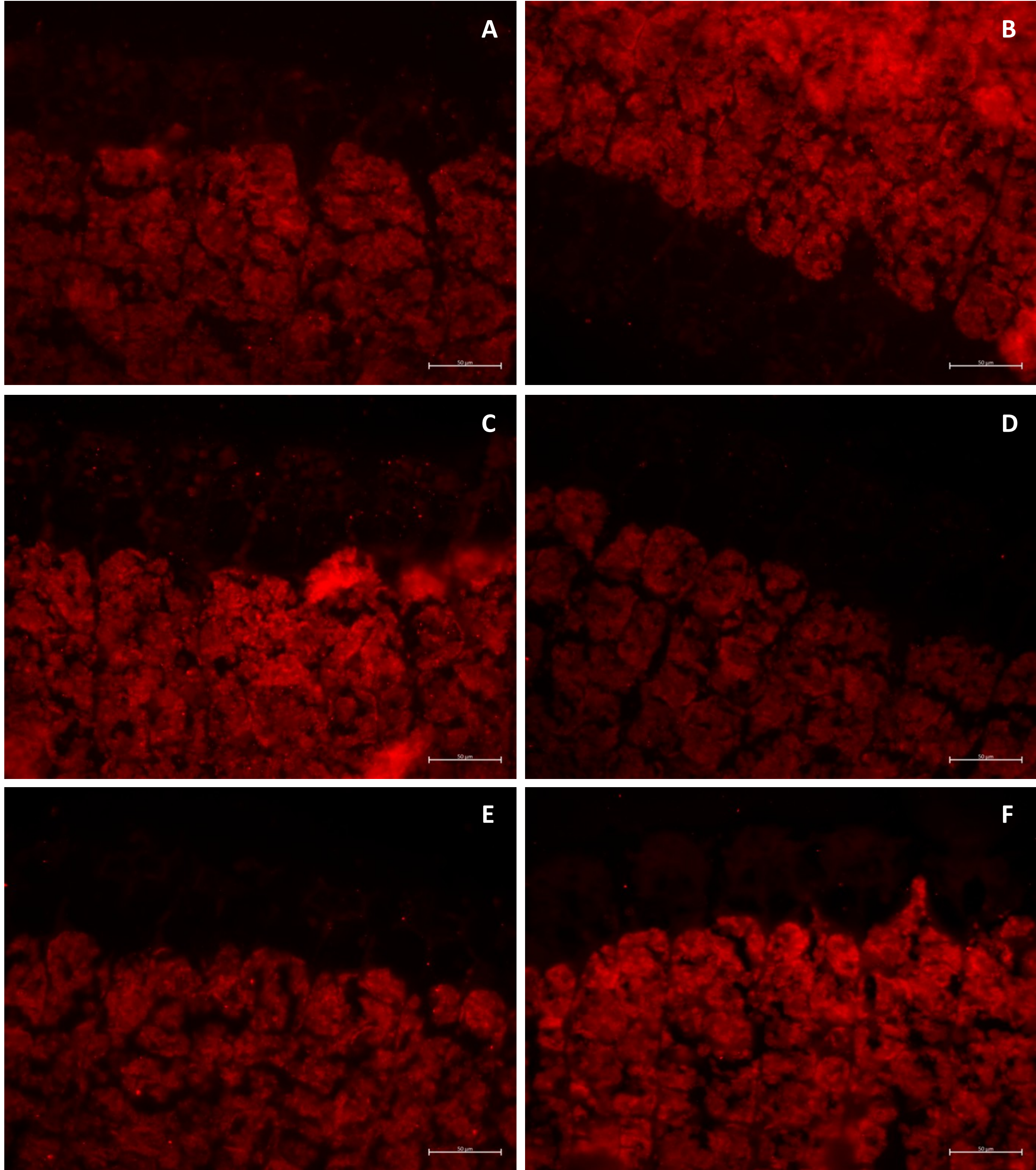
All oligonucleotide probes were used in a working solution concentration of 50 ng/µl.

1. Prepare Hybridization buffer and probe solution: Add 1,5 µl of probe working solution to 20 µl of hybridization buffer for each different probe mix
2. Put a kimwipe into a 50 ml Falcon tube to form a humid chamber
3. Add 2 ml of hybridization buffer, soak well and stuff it down to the bottom of the tube
4. Form rings with a Pap-Pen around each section on the microscope slide to separate the individual sections during hybridization
5. Pipette 20 µl of each respective hybridization buffer-probe solution onto every section
6. Put the slide horizontally into the tube and incubate in this position at 46°C for 3 hours
7. After 3 hours remove hybdidization buffer from the section by pushing the slide shortly sideways onto the table
8. Wash the slide in one tube with washing buffer. Then put the slides into another tube with washing buffer and incubate at 48°C for 15 minutes
9. Remove slides from tube and was quickly in ice cold ddH_2O. Filter sections that have fallen off into the tube can be washed in ddH_2O individually.
10. Dry slides with compressed air. Filter sections have to be dried individually and placed back on the slide
11. Pipet 20 µl of DAPI solution onto the sections and incubate for 15 minutes
12. After 15 minutes wash sections with ice cold ddH_2O and dry with compressed air

13. Add one drop of Citifluor antifadent onto every section and cover the slides with a coverslip
14. Seal with nail-polish and store slides at 4°C in the dark

12.2 Supplementary Figures

12.2.1 Formamide Series



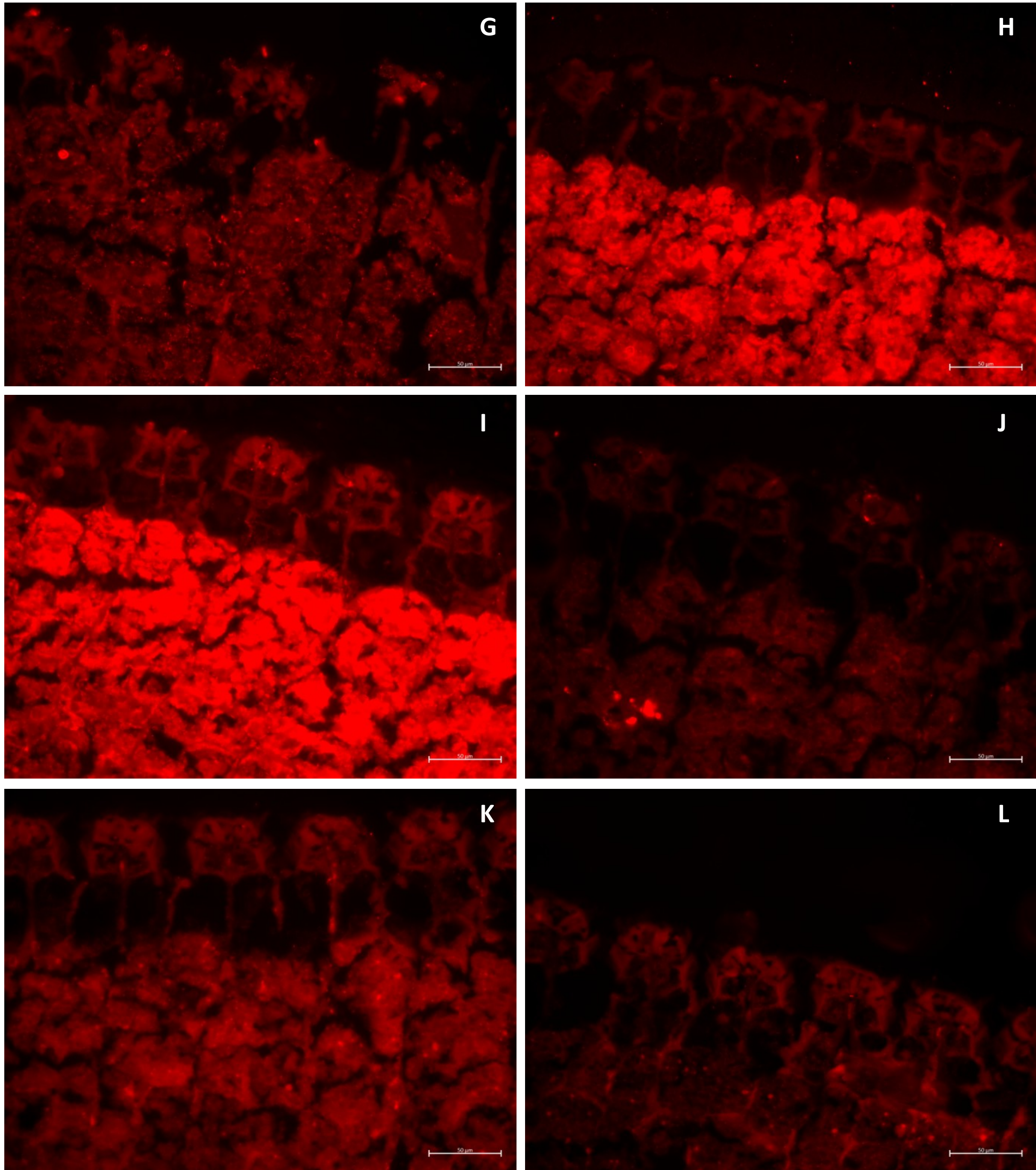


Figure 15: Fluorescence microscopy pictures of *Loripes orbiculatus* gills using the 845-Cy3 oligonucleotide probe with different formamide concentrations. Channel: Rhodamine B (emission: 575 – 640nm). Scalebars: 50µm
A) 0%; **B)** 5%; **C)** 10%; **D)** 15%; **E)** 20%; **F)** 25% **G)** 30% **H)** 35% **I)** 40% **J)** 50% **K)** 60% **L)** negative control with non-338-Cy3 probe. The intensity of the fluorescence signal indicates the rate of probe binding. Probe 845 seems to bind best at 35 – 40% formamide concentration in the hybridization buffer.