December 2018

Taxonomic, Genetic and Functional Diversity of Symbionts Associated with the Coastal Bivalve Family Lucinidae

Jean S. Lim
Clemson University, shenjean@gmail.com

TAXONOMIC, GENETIC AND FUNCTIONAL DIVERSITY OF SYMBIONTS ASSOCIATED WITH THE COASTAL BIVALVE FAMILY LUCINIDAE

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Biological Sciences

---

by
Lim Shen Jean
December 2018

Accepted by:
Barbara J. Campbell, Committee Chair
Antonio J. Baeza
Annette S. Engel
Vincent P. Richards

ABSTRACT

Extant bivalve members from the family Lucinidae harbor chemosynthetic gammaproteobacterial gill endosymbionts capable of thioautotrophy. These endosymbionts are environmentally acquired and belong to a paraphyletic group distantly related to other marine chemosymbionts. In coastal habitats, lucinid chemosymbionts participate in facilitative interactions with their hosts and surrounding seagrass habitat that results in symbiotic sulfide detoxification, oxygen release from seagrass roots, carbon fixation, and/or symbiotic nitrogen fixation. Currently, the structural and functional complexity of whole lucinid gill microbiomes, as well as their interactions with lucinid bivalves and their surrounding environment, have not been comprehensively characterized. This dissertation focuses on the taxonomic, genetic, and functional diversity in the gill microbiomes of three Floridian coastal lucinid bivalve species, *Phacoides pectinatus*, *Ctena orbiculata*, and *Stewartia floridana*, in the context of environmental data where appropriate.

Analyses of these lucinid gill microbiomes showed taxonomic diversity that was unaffected by spatial distribution patterns. *Phacoides pectinatus* gill microbiomes sampled from a coastal mangrove habitat contained, in order of relative abundances, a chemosynthetic symbiont species that was taxonomically and functionally distinct from seagrass-associated chemosynthetic lucinid symbionts, a heterotrophic *Kistimonas*-like species, and a heterotrophic *Spirochaeta*-like species. In comparison, gill microbiomes of a seagrass-dwelling *C. orbiculata* population comprised four strains of chemosymbionts that belonged to two separate species and low abundances of an uncharacterized

Endozoicomonas-like operational taxonomy unit (OTU). Gill microbiomes of a separate seagrass-dwelling *S. floridana* population consisted of another chemosynthetic symbiont species and low abundances of a heterotrophic *Spirochaeta*-like species that was distantly related to the *Spirochaeta*-like species in *P. pectinatus*.

Functional characterization of host- and microbiome-related genes/transcripts in these bivalve species revealed previously unreported C1-compound oxidation functions in some chemosymbionts and other functions relevant to microbe-microbe competition, symbiont selection, metabolism support, and symbiont-to-host nutrient transfer. Preliminary differential expression analyses on host- and microbiome genes across micro-habitats with different vegetation coverages showed potential upregulation of *C. orbiculata* functions involved in aerobic respiration, aerobic stress, electron transport, and mitochondrial sulfide detoxification, as well as downregulation of a sulfurtransferase gene encoded by its chemosynthetic symbionts, in a seagrass-covered quadrat compared to an algae-covered quadrat. In comparison, very few genes mappable to *S. floridana* and its chemosymbiont were differentially expressed between predominantly sand-covered and seagrass-covered quadrats, but the *Spirochaeta*-like species over-expressed carbon, nitrogen, phosphate, transport, synthesis, transcriptional regulation, and protein degradation functions in predominantly sand-covered quadrats.

These findings reaffirm the overlooked notion of heterogeneous lucinid gill microbiomes that can vary within and between host species and populations. At the same time, this project advances understanding of the functional diversity across chemosynthetic lucinid symbionts and offers insights on lucinid-microbiome-environment interactions.

DEDICATION

This thesis is dedicated to my beloved dog Blackie and my grandmother whom, with much regret and sadness to this day, I was unable to bid farewell to.

ACKNOWLEDGMENTS

I am touched by the unwavering faith, understanding, and support of my dearest family and friends as I go through this Ph.D. program. I would like to thank my advisor, Dr. Barbara Campbell, committee members, as well as my fellow graduate students and staff in the Department of Biological Sciences for their guidance and advice on research, administrative processes, and life in general. Specifically, I would like to thank Dr. Vincent Richards and his former graduate students, Abdullah Abood and Yirui Chen, for their time and effort spent helping me with MiSeq sequencing. Beyond my department, I am grateful to the Clemson University Palmetto Cluster; Barbara Blackmon and Michael Atkins from the now-defunct Clemson University Genomics Institute (CUGI); Dr. Terri Bruce and Rhonda Powell from the Clemson University Light Imaging Facility (CLIF); Chad McMahan from the Biomaterials/Histology lab in the Clemson Bioengineering Department, whom or which have greatly facilitated my research activities. I would also like to thank the Clemson University Creative Inquiry program for funding undergraduate research that has enriched my research experience and benefited my dissertation work. Finally, I am grateful to my collaborators, including Dr. Annette S. Engel, Audrey Paterson, and their lab colleagues at the University of Tennessee at Knoxville, as well as Dr. Laurie C. Anderson, Brooke Long, Broc Kokesh, and their lab colleagues at the South Dakota School of Mines and Technology for their contributions to field work, data analyses, ideas, and discussions on the project. I would also like to thank the National Science Foundation's Dimensions of Biodiversity Program (DEB-1342763) for funding my research project.

TABLE OF CONTENTS

Table of Contents (Continued)

Page

LIST OF TABLES

List of Tables (Continued)

Table Page

.

LIST OF FIGURES

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

CHAPTER ONE

INTRODUCTION

Evolutionary history of the Lucinidae

The Lucinidae family of clams (lucinids) belong to the kingdom Animalia, phylum Mollusca and class Bivalvia. They represent an archaic bivalve clade with a long fossil record. Arguably the oldest lucinid fossil, *Ilionia prisca*, dates back to the Silurian period within the Paleozoic era (Liljedahl, 1992). *Ilionia prisca* possessed morphological traits similar to modern lucinids, such as anterior shell extension, gut reduction, the separation of the elongated anterior adductor muscle from the pallial line, a diagonal sulci suggestive of the existence of a posterior exhalant siphon, and the absence of an inhalant siphon (Liljedahl, 1992). The life position, and consequently, life habits, of *I. prisca* were deduced to be similar to modern lucinids, because fossils of the former were discovered in deep sediments with their anterior sides oriented at an angle against the direction of the waves (Liljedahl, 1992). Morphological evidence, as well as the presence of pyrite hinting at low oxygen concentrations and high sulfur concentrations in its habitat, suggests that chemosymbiosis in *I. prisca* was possible (Liljedahl, 1992). After the Silurian period, coastal lucinids existed at a low diversity before undergoing a diversification burst during the late Cretaceous period (Stanley, 2014). This evolutionary radiation was attributed to the emergence of seagrasses and mangroves, whose sulfide-rich sediments sustain the

growth of free-living thioautotrophic bacteria and thus provide lucinids with a supply of potential symbionts (Stanley, 2014).

To accommodate their gill endosymbionts, lucinid bivalves have undergone a series of morphological adaptations through evolutionary time. Because of increasingly intimate associations and possible co-evolution with free-living chemosynthetic bacteria, lucinid bivalves developed specialized epithelial symbiont-containing cells, known as bacteriocytes, that allowed bacteria to move from an extracellular location to an intracellular position in the gills (Taylor, 2000). As the gills simplified and enlarged to house more bacteriocytes, the outer demibranch was lost (Taylor, 2000). The gills possibly displaced the main respiratory channel, which is now an inner mantle around the anterior adductor muscle thickened by blood space or transformed into mantle gills (Taylor, 2000; Taylor and Glover, 2009; Figure 1.1). In addition, the anterior adductor muscle in many lucinid species is elongated and detached from the pallial line, which separates the anterior and posterior cavities and prevents sulfide oxidation as oxygenated water flows through these cavities (Taylor, 2000; Taylor and Glover, 2009; Figure 1.1). With growing dependence on chemosymbiosis, the anterior inhalant tube became the main water conducting channel as the posterior inhalant opening lost its function (Reid, 1986; Figure 1.1). Consequently, lucinids reorganized their life position, where the anterior, instead of the posterior side, became angled upward towards the water column for efficient water and nutrient uptake (Taylor, 2000; Roeselers and Newton, 2012; Anderson 2014). Unlike other bivalves, lucinids lack an inhalant siphon, possibly because it interfered with sulfide uptake (Reid, 1986; Liljedahl, 1992). Instead, their vermiform, extensible foot excavates ventral

tunnels for porewater sulfide acquisition (Liljedahl, 1992; Taylor, 2000; Taylor, 2010; Figure 1.1). The foot also constructs an anterior inhalant tube for the transport of water containing food particles and oxygen to the mouth and gill, respectively (Liljedahl, 1992; Taylor, 2000; Taylor, 2010; Figure 1.1). The posterior part of the foot facilitates movement and burrowing (Liljedahl, 1992; Taylor, 2000; Figure 1.1). Digestive system adaptations, including the reduction of the labial palps, stomach, and gut, are also observed in some lucinid species, such as *Loripes orbiculatus*, but not in other species, such as *Lucinella divaricata* (Le Pennec *et al.*, 1995; Taylor, 2000; see also "Particulate feeding in lucinid clams" subsection below).

Today, more than 100 living lucinid species from approximately 69 genera are recognized (NCBI Resource Coordinators, 2016; Taylor *et al.*, 2016). Based on their 18S rRNA gene, 28S rRNA gene, and cytochrome b gene phylogenies, lucinids are classified into five major subfamilies (Codakiinae, Leucosphaerinae, Lucininae, Myrteinae, and Pegophyseminae) and two minor subfamilies (Fimbriinae and Monitilorinae; Taylor *et al.*, 2016). Among these, the Lucininae subfamily consists of >43 extant genera and the highest number of species, whereas the Fimbriinae and Monitilorinae subfamilies each consist of a single living species, *Fimbria fimbriata* and *Monitilora ramsayi*, respectively (Taylor *et al.*, 2016).

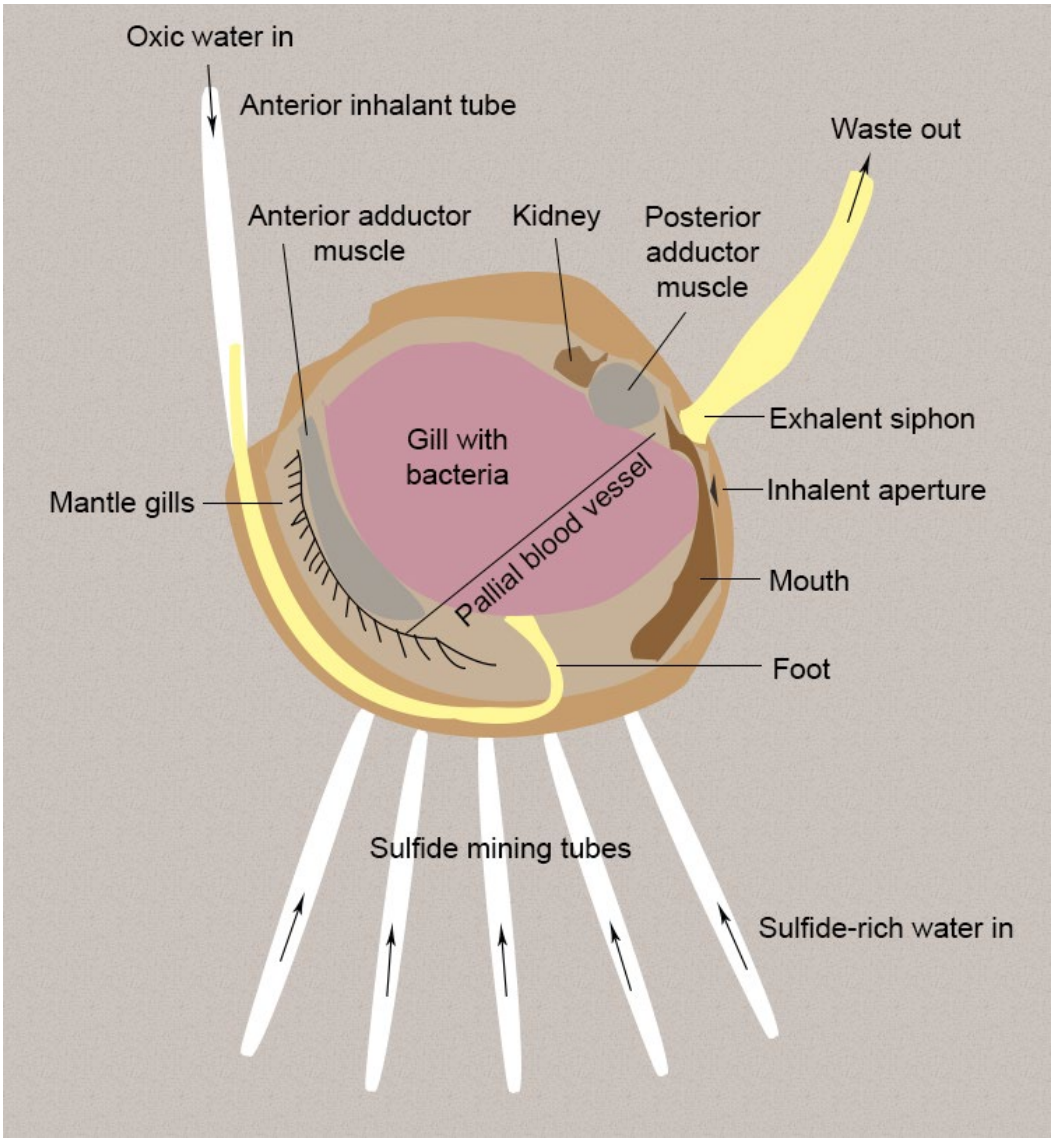


Figure 1.1. The general morphology, life position, and water flow of a lucinid bivalve based on *Phacoides pectinatus* and *Codakia* spp., modified from Taylor and Glover, 2000.

Particulate feeding in lucinid clams

In type 3 reducing-system chemosymbiotic bivalves that include lucinid clams, digestive system phenotypes are not commonly shared and, instead, correspond to levels of sulfide in their habitats and pallial cavities (Le Pennec *et al.*, 1995; Le Pennec and Beninger, 2000). Increasing sulfide concentrations were thought to lead to progressive digestive system adaptations ranging from labial palp reduction (in conjunction with symbiotic association), intestinal reduction (in conjunction with gill hypertrophy), stomach reduction, loss of crystalline style and style sac, digestive tubule reduction, loss of digestive glands to the complete loss of the digestive system (Le Pennec *et al.*, 1995).

Despite inter-species differences in digestive system morphology and physiology, many extant lucinid species including *Codakia orbicularis*, *Divaricella quadrisulcata*, *Linga pensylvanica*, *Loripes orbiculatus*, *Lucinella divaricata*, *Lucinoma aequizonata*, *Lucinoma borealis*, *Parvilucina tenuisculpta*, *Phacoides pectinatus*, and *Stewartia floridana* are capable of particulate feeding (Le Pennec *et al.*, 1995; Le Pennec and Beninger, 2000; Duplessis *et al.*, 2004a; van der Geest *et al.*, 2014). Heterotrophy in these species were inferred from the presence of particulate organic matter and/or digestive enzymes in their digestive systems, gill morphology, radiolabel ingestion experiments, and slightly enriched $\delta^{13}C$ ratios in non-gill tissues compared to gill tissues or thioautotrophic symbiont bacterial pellets free from host contamination (Le Pennec *et al.*, 1995; Duplessis *et al.*, 2004a; Rossi *et al.*, 2013; van der Geest *et al.*, 2014; see also "The lucinid-bacteria chemosymbiosis" subsection below). Besides particulate matter, dissolved free amino acids have also been postulated to serve as a carbon source in a *Lucinoma aequizonata*

population at the Santa Barbara Basin, USA (Cary *et al.*, 1989). Lucinid bivalves feed themselves by filtering seawater containing particulate matter or dissolved substrates through anterior mucus-lined inhalant tubes excavated by their feet at the sediment-water interface (Taylor, 2000). Filtered water is propelled by the beating of cilia on the gills, where particulate matter is trapped in gill mucociliary epithelium and transported to the mouth next to the foot (Taylor, 2000; Duplessis *et al.*, 2004a).

The lucinid-bacteria chemosymbiosis

To supplement particulate feeding (see also "Particulate feeding in lucinid clams" subsection above), obligate intracellular chemosymbionts fulfill some or most of a host's nutritional needs (Spiro *et al.*, 1986; van der Geest *et al.*, 2014). Chemosymbiosis was first described in *Riftia pachyptila*, a hydrothermal vent tube worm (Cavanaugh *et al.*, 1981; Felbeck, 1981). In the trophosome tissue of *R. pachyptila*, high activities of rhodanese, adenosine 5′-phosphosulfate-reductase (Apr), and adenosine triphosphate (ATP)-sulfurylase (Sat) involved in sulfur oxidation, as well as ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and ribulose 5-phosphate kinase (Ru5P kinase) involved in carbon fixation, were detected by enzymatic assays (Felbeck, 1981). Visualization of the trophosome tissue of using electron and epifluorescence microscopy revealed the presence of sulfur granules and the presence of the bacterial outer cell wall component lipopolysaccharide (LPS) (Cavanaugh *et al.*, 1981). Since then, similar discoveries of thioautotrophic symbionts were reported in bivalve species inhabiting sulfide-rich habitats, including *Solemya velum* (Cavanaugh, 1983), *Calyptogena*

magnifica, *Calyptogena pacifica* from deep-sea vent sites, as well as *Solemya panamensis,* and the lucinid bivalves *Parvilucina tenuisculpta* and *Lucinoma annulata* from coastal environments (Felbeck *et al.*, 1981).

In lucinids, endosymbionts occur in bacteriocyte vacuoles, where one to two individuals occupy most of the vacuole volume based on electron microscopy, (Fisher and Hand, 1984; Reid and Brand, 1986; Johnson and Fernandez, 2001; Ball *et al.*, 2009). Symbiont thioautotrophic functions have been verified through histological methods (e.g., light microscopy, transmission electron microscopy and epifluorescent microscopy, X-ray), LPS detection assays, and enzymatic assays (Felbeck *et al.*, 1981; Fisher and Hand, 1984; Dando *et al.*, 1985; Schweimanns and Felbeck, 1985; Spiro *et al.*, 1986), and $\delta^{13}C$ analyses confirm that lucinids are nutritionally dependent on symbiotic carbon fixation (Spiro *et al.*, 1986; van der Geest *et al.*, 2014). Lucinid and other chemosymbiotic bivalves have $\delta^{13}C$ values that range from -23 per mil ‰ (*Lucinisca nassula*, formerly *Lucina nassula*) to -31 ‰ (*Thyasira sarsi;* Spiro *et al.*, 1986; Duperron *et al.*, 2007; Rodrigues *et al.*, 2010; Duperron *et al.*, 2012; van der Geest *et al.*, 2014) because the enzyme RuBisCO prefers ^{12}C over ^{13}C when fixing CO_2 from porewater bicarbonate or respiration waste products (Spiro *et al.*, 1986). In contrast, tissues of heterotrophic marine invertebrates exhibit slight enrichment in $\delta^{13}C$ values compared to their food sources, possibly because ^{12}C is preferentially lost during respiration, ^{13}C is preferred during carbon assimilation, or enzymatic biochemical reactions possess intrinsic ^{13}C-enrichment properties (Spiro *et al.*, 1986). One histoautoradiographic study of *Loripes orbiculatus* shows that some fixed CO_2 can also be translocated to other symbiont-free tissues (Herry *et al.*, 1989).

Besides thioautotrophy, other metabolic functions have also been described in chemosynthetic lucinid endosymbionts. Results from incubation experiments, oxygen, nitrate and nitrite measurements, nitrate reductase assays, and/or 5-cyano-2,3-ditolyl tetrazolium chloride assays, determined the endosymbiont from *Codakia orbicularis* can respire solely on oxygen (Duplessis *et al.*, 2004a), but *Lucinoma aequizonata* can respire solely on nitrate (Hentschel *et al.*, 1993; Hentschel and Felbeck, 1995; Hentschel *et al.*, 1996); *Ctena orbiculata* can co-respire or respire alternately on both oxygen or nitrate (Barnes, 1993). Activity of a symbiont-related ferredoxin-dependent nitrite reductase, which converts nitrite to ammonia, was first recorded in the gills of the lucinid *Stewartia floridana* (formerly *Lucina floridana*) (Fisher and Hand, 1984). Consistent with previous observations, recent genomic analyses on the chemosynthetic symbionts in *C. orbicularis* and *L. orbiculatus* have identified symbiotic genes associated with aerobic respiration and assimilatory and/or dissimilatory denitrification (König *et al.*, 2016; Petersen *et al.*, 2016). These -omics-centered studies have also revealed mixotrophy and hydrogen oxidation symbiotic functions previously characterized in related chemosymbiotic marine organisms (Woyke *et al.*, 2006; Petersen *et al.*, 2011; Dmytrenko *et al.*, 2014; Nakagawa *et al.*, 2014; Kleiner *et al.*, 2015; see also "Phylogeny of chemosynthetic lucinid symbionts" subsection below), but not yet reported in lucinid bivalves, as well as novel functions previously undiscovered for chemosymbiotic marine organisms, including urea decomposition and/or nitrogen fixation (König *et al.*, 2016; Petersen *et al.*, 2016).

To sustain the lucinid-bacteria chemosymbiosis, some form of host-to-symbiont transport and metabolic transfer is necessary, including involvement of hemoglobins,

respiratory pigments, peroxisomes, lysosomes, and other features. Specifically, similar to the hydrothermal vent tube worms *Ridgeia piscesa* and *R. pachyptila* (Arp and Childress, 1981; Arp and Childress, 1983; Carney *et al.*, 2007), the lucinid bivalve *Phacoides pectinatus* produces high levels of hemoglobins 1, 2, and 3 that can transport sulfide and oxygen to their gill symbionts (Kraus and Wittenberg, 1990; Frenkiel *et al.*, 1996; Rizzi *et al.*, 1996). In contrast, hemoglobins binding oxygen, but not sulfide, were identified in *Myrtea spinifera* (Dando *et al.*, 1985). The presence of uncharacterized hemoglobin was also reported in *Myrtea flabelliformis* (Brissac *et al.*, 2011), while hemoglobin-like cells were visualized in *Anodontia ovum* (Ball *et al.*, 2009). Iron was also detected in the gill pigment granules of *S. floridana*, hinting at the possible presence of respiratory pigments, such as hemoglobin and myoglobin and/or Fe-containing cytochromes (Fisher and Hand, 1984). Nevertheless, besides *P. pectinatus*, hemoglobin types and functions in other lucinid species have not been studied in detail. Peroxisomes, visualized in gill bacteriocytes of *Linga pensylvanica*, have been hypothesized to prevent spontaneous sulfide oxidation in the gill or perform oxidation reactions related to symbiotic thiotrophy (Gros *et al.*, 1996a). Sulfide-oxidizing bodies presumably involved in host-mediated sulfide detoxification have also been visualized in gill bacteriocytes of *P. pectinatus* (Liberge *et al.*, 2001). Additionally, the foot of *L. aequizonata*, which has higher sulfide oxidase activity than the mantle, has been postulated to partially oxidize sulfide to thiosulfate that can be used by chemosynthetic gill symbionts as a possible means of reducing sulfide toxicity in the blood (Cary *et al.*, 1989). Albeit producing less energy than sulfide, thiosulfate is non-toxic and less-diffusible, allowing it to be concentrated in the blood and transported to the

chemosynthetic gill symbionts together with low amounts of sulfide (Cary *et al.*, 1989). Lastly, lucinid hosts may acquire nutrition from their chemosynthetic symbionts through enzymatic lysis because lysosomes and/or associated hydrolases, such as acid phosphatase and/or arylsulfatase, have been detected within gill bacteriocytes of *Anodontia ovum*, *C. orbicularis*, *Lucina* (formerly *Linga*) *pensylvanica*, *Loripes orbiculatus*, and *P. pectinatus* (Frenkiel *et al.*, 1996; Oliver *et. al.*, 1996; Johnson and Fernandez, 2001; Liberge *et al.*, 2001; Ball *et al.*, 2009). Glycogen particles abundant within symbiont cells of *L. orbiculatus* could potentially serve as a carbon source for the host (Johnson and Fernandez, 2001).

Lucinid growth and reproduction

In general, bivalves grow with accretionary increments in their shell sizes, as $CaCO_3$ is deposited at the mantle edge and the shell within the pallial line thickens simultaneously (Jones and Quitmyer, 1996). This growth process is observable through annual growth bands external and internal to the shell and is affected by environmental variations, such as rhythmic diurnal, tidal and seasonal cycles (Jones and Quitmyer, 1996). Based on shell growth patterns, a lifespan of two to three years was estimated for some type 3 reducing-system chemosymbiotic bivalves. In *S. floridana*, a maximum lifespan of six years was deduced (Long, 2016).

Reproduction in lucinid bivalves generally occurs in summer months as water temperature rises. Lucinid species including *C. orbicularis*, *L. orbiculatus* and *P. pectinatus*, are mostly dioecious, though small numbers of hermaphrodite *C. orbicularis*

clams and sequential hermaphroditism in *P. pectinatus* have been reported (Alatalo *et al.*, 1984; Berg and Alatalo, 1984; Le Pennec and Beninger, 2000; Christo *et al.*, 2016). Histological observations of gamete presence and maturity suggested discontinuous gametogenesis in *C. orbicularis* and *L. orbiculatus*, (Le Pennec and Beninger, 2000). *Codakia orbicularis* observed between 1981-1982 from Gold Rock Creek, Bahamas developed their gonads in spring as water temperature increased and did not undergo gametogenesis between August to March (Alatalo *et al.*, 1984). In comparison, *L. orbiculatus* reproduces semi-annually. In an intertidal *L. orbiculatus* population off the coast of Mauritiania, increased gametogenesis occurred in March-July and September-February, and major spawns occurred in January-February and July-August between 2009-2011 (van der Geest *et al.*, 2014). In another *L. orbiculatus* population at Brittany, France, major and minor spawns took place in May and November-December 1991, respectively (Johnson and Fernandez, 2001). Growth and gametogenesis in both *L. orbiculatus* populations appeared to require increased levels of nutrition derived from particulate feeding (Mauritiania population) or symbiont lysis (French population; Chapter I "The lucinid-bacteria chemosymbiosis" subsection; Johnson and Fernandez, 2001; van der Geest *et al.*, 2014). On the other hand, although discontinuous gametogenesis was similarly inferred in *P. pectinatus* based on variability in the thickness of the species' oocyte-adhering acinal wall, the low numbers and consistent maturity of their oocytes, together with the high numbers of (≥40%) of sampled individuals from the Paranaguá Estuarine Complex, Brazil with partially filled gonads between January-March (summer/autumn) and June-August (winter) 2009, suggested a continuous reproductive cycle (Le Pennec and

Beninger, 2000; Christo *et al.*, 2016). Major spawning of the Paranaguá *P. pectinatus* population occurred in the summer, where increased gonad repletion was observed with rising water temperatures (Christo *et al.*, 2016).

Embryonic development events after spawning have been well-characterized in *C. orbicularis* (Gros *et al.*, 1997). In this species, fertilized eggs undergo cell division and develop into swimming ciliated trochophores within 24 hours (Gros *et al.*, 1997). 39 hours after fertilization, trochophores develop into ciliated veligers with symmetrically differentiated organs along both sides of the mouth-anus axis and enlarged calcified shells that mark the final larval stage (Gros *et al.*, 1997). Straight-hinged (D-shaped), ciliated swimming veligers hatch from egg capsules 48 hours post-fertilization and continue to feed on egg-derived vitelline until after the first week (Gros *et al.*, 1997). 15 days post-fertilization, veligers develop functional foot and become pediveligers capable of swimming and crawling (Gros *et al.*, 1997). One day later, pediveligers lose their swimming functions and develop into crawling benthic plantigrades as the ciliary velum regresses rapidly (Gros *et al.*, 1997). Metamorphosis occurs five weeks post-fertilization and involves differentiation of the gill filaments, byssal gland (clear cells within and along the heel to the tip of the foot), exhalant siphon, and secretion of a fast-growing modified shell (dissoconch; Gros *et al.*, 1997).

Lucinid habitats

The lucinid-bacteria chemosymbiotic association allows lucinids to colonize habitats scarce in food and low in oxygen and high in sulfide, such as oxic-anoxic interfaces in sediments of diverse marine environments (Liljedahl, 1992; Durand and Gros, 1996; Taylor and Glover, 2000; Taylor and Glover, 2010). These habitats are otherwise uninhabitable by other bivalves, and often have low bivalve-species diversity (Liljedahl, 1992). Consequently, compared to other bivalves, lucinids occupy the widest range of habitats and geographical locations (Roeselers and Newton, 2012). Lucinids have been documented in oxygen minimum zones (*Lucinoma* spp.; Cary *et al.*, 1989; Oliver and Holmes, 2006), deep sea sediments (*Gonimyrtea ferruginea* and *Myrtina reflexa*; Taylor and Glover, 2013), cold seeps (*Lucinoma aff. kazani*; Duperron *et al.*, 2007), mud volcanoes (*Lucinoma* spp.; Rodrigues *et al.*, 2010), and in hydrothermal vents (*Bathyaustriella thionipta*; Glover *et al.*, 2004). In shallow marine environments, lucinids have been recorded in organic-rich, reducing sediments around mangroves swamps (*Anodontia* spp., *Austriella corrugata*, *Indoaustriella* spp., *Pillucina vietnamica*, *Phacoides pectinatus*; Frenkiel *et al.*, 1996; Primavera *et al.*, 2002; Glover *et al.*, 2008; Meyer *et al.*, 2008;), a sewage outfall (*Loripes orbiculatus*; Herry *et al.*, 1989), an intertidal mud flat (*Lucinoma borealis*; Dando *et al.*, 1994), and most frequently in tropical and temperate seagrass beds (e.g., Fisher and Hand, 1984; Barnes, 1993; Johnson *et al.*, 2002; Green-García and Engel, 2012; Rossi *et al.*, 2013; Taylor and Glover, 2013; Reynolds *et al.*, 2014; Sanmartí *et. al.*, 2017). Predators of bivalves in shallow marine environments include fishes such as the big-scaled sand smelt *Atherina boyeri*, the common goby

Pomastochistus microps, the gilthead seabream *Sparus aurata* (Rossi *et al.*, 2013), as well as the spiny Caribbean lobster *Panulirus argus* (Higgs *et al.*, 2016).

Lucinid bivalves inhabit sediments with varying nutrient concentrations. They have been discovered in sub-oxic sediment cores with <20 μM dissolved oxygen concentrations and possibly oxic sediments with porewater oxygen concentrations in the millimolar range (Table 1.1). Recorded porewater nitrate and nitrite concentrations of lucinid habitats range from <32 μM to 0.1 mM and ≤1 μM, respectively, whereas ammonium concentrations have increasing (Cary *et al.*, 1989) or decreasing trends (Barnes, 1993) with sediment depth (Table 1.1). Sediments harboring lucinid species range from organic-poor (0.5% organic carbon content) to organic-rich (>3% organic carbon content; Table 1.1). Thiosulfate concentrations in interstitial waters have been reported to decrease in sediment depth (Barnes, 1993) or are detectable only in some sections of ≤ 11 cm depth sediment cores (Cary *et al.*, 1989; Table 1.1). Porewater sulfate concentrations at lucinid sampling sites are between 30 mM to 38 mM, which reflect the marine habitat conditions, and porewater sulfide concentrations range from nanomolar to millimolar concentrations (Table 1.1). Other environmental variables, such as air temperature, water temperature, water pH, oxidation-reduction potential, sediment type, grain size, seagrass species, acid-volatile sulfide concentrations, metal concentrations, carbon dioxide/monoxide concentrations, hydrogen concentrations, and methane concentrations have been described but are not as comprehensively reviewed to understand how these parameters influence lucinid diversity and growth (Dando *et al.*, 1986; Cary *et al.*, 1989; Barnes, 1993; Hentschel *et al.*, 1993; Reynolds *et al.*, 2007; Meyer *et al.*, 2008; Green-García and Engel, 2012).

Table 1.1. Lucinid species and a subset of nutrient concentrations measured in their sampling sites. Data values are obtained from literature. "ND" indicates no data.

Lucinid species	*Lucinoma borealis*	*Lucinoma aequizonata*		*Ctena orbiculata*	*C. orbiculata* & *Lucinisca nassula*	*Pillucina vietnamica* & *Indoaustriella dalli*	*Lucinisca nassula*
Study site	Mill Bay, England	Santa Barbara Basin, California, USA		Bailey's Bay, Bermuda	Florida Bay, Florida, USA	Kungkrabaen Bay, Thailand	Cedar Key, Florida, USA
Oxygen	ND	18-20 µM	4-7 µM	101-119 %	ND	ND	0.06 mM
Dissolved organic carbon	ND	3-5 %	ND	5-11%	ND	0.5-3%	2.5-1.4%
Nitrate	ND	<25 µM	29-32 µM	0.4-6 µM	ND	ND	0.1 mM
Nitrite	ND	ND	0.01-1 µM	<1 µM	ND	ND	ND
Ammonium	ND	<10-<250 µM	ND	5-31 µM	>5-<25 µM	ND	ND
Thiosulfate	≤300 nM	<2->6 µM (if detectable)	ND	0.06-32 µM	ND	ND	ND
Sulfate	ND	~30 mM	ND	ND	ND	ND	38 mM
Sulfide	≤200 nM	>3->8 µM (if detectable)	ND	11-394 µM	>60-<80 µM	ND	0.05 mM
Reference	Dando *et al.*, 1986	Cary *et al.*, 1989	Hentschel *et al.*, 1993	Barnes, 1993	Reynolds *et al.*, 2007	Meyer *et al.*, 2008	Green-García and Engel, 2012

Lucinid-bacteria-environment interactions

The frequent associations of lucinid clams with seagrass habitats have led to the development of a three-partner symbiosis model that involves facilitative interactions between lucinids, their chemosynthetic symbionts, and their surrounding seagrass beds (van der Heide *et al.*, 2012). In seagrass sediments, sulfate-reducing microorganisms decompose dead organic matter and release potententially phytotoxic levels of hydrogen sulfide gas (Reynolds *et al.*, 2014; van der Geest *et al.*, 2014). Chemosynthetic lucinid symbionts oxidize reduced sulfur compounds and fix carbon, as well as potentially nitrogen, within the sediments, which removes sulfide and generates organic nitrogen that can promote seagrass growth (Johnson *et al.*, 2002; van der Heide *et al.*, 2012; Reynolds *et al.*, 2014; Petersen *et al.*, 2016). In return, seagrass roots supply oxygen to support the respiratory needs of the lucinid bivalve hosts and their symbionts (Fisher and Hand, 1984; van der Heide *et al.*, 2012; Sanmartí *et. al.*, 2017).

In loose agreement with the tripartite symbiosis model, abundances of lucinid bivalves have been positively associated with the presence and/or root complexity of seagrass beds (Fisher and Hand, 1984; Sanmartí *et. al.*, 2017; also see "Lucinid habitats" subsection, above). Based on measured RuBisCO activities, net production of lucinid bivalves in seagrass beds was estimated to be 0.003 (for *Myrtea spinifera*) to 2 (for *Stewartia floridana*) grams of carbon per m^2 per year (Fisher and Hand, 1984; Dando *et al.*, 1985; Johnson *et al.*, 2002). Reynolds *et al.* (2014) also provide mechanistic evidence for the tripartite symbiosis model by demonstrating that, compared to sampling sites without lucinid species, significantly lower sulfide and higher ammonium porewater

concentrations were measured at Florida Bay in sampling sites containing *C. orbiculata* and *Lucinisca nassula* (Reynolds *et al.*, 2014). The authors of the same study estimated lucinid-mediated sulfide removal to be 2-16% of the total sulfide produced in seagrass sediments and additionally demonstrated that *C. orbiculata* and *L. nassula* significantly reduced porewater sulfide concentrations in a microcosm experiment (Reynolds *et al.*, 2014). However, beyond these studies, investigations of the tripartite symbiosis model are scarce.

An overview of symbiont transmission modes

Symbiont transmission plays a crucial role in the establishment, maintenance, and evolution of symbiosis (Bright and Bulgheresi, 2010). Essentially, there are two symbiont transmission modes. Horizontal transmission entails symbiont acquisition from an environmental, free-living stock, and vertical transmission involves the transmission of symbionts through host (usually female) gametes (Bright and Bulgheresi, 2010). A mixed transmission mode also exists, where either environmental transfer or intra-species and inter-species symbiont transfer occur in hosts with vertically transmitted symbionts (Bright and Bulgheresi, 2010).

Vertical transmission is generally believed to drive the evolution of symbiosis (Genkai-Kato and Yamamura, 1999; Bright and Bulgheresi, 2010). Vertically transmitted symbionts undergo population bottlenecks during initial colonization and subsequent transmissions, which leads to increased genetic drift (Dubilier *et al.*, 2008). Under the prediction of the nearly neutral theory of molecular evolution, increased genetic drift,

coupled with weak selection pressure, leads to accelerated mutation and fixation rates of near-neutral and deleterious alleles (Peek *et al.*, 1998; Stewart and Cavanaugh, 2006; Dubilier *et al.*, 2008). This eventually results in gene loss and reduction in genome sizes of vertically transmitted symbionts (Stewart and Cavanaugh, 2006; Dubilier *et al.*, 2008). In contrast, strong selection pressure would have removed highly deleterious mutations from the population and thus is not included in the nearly neutral theory of molecular evolution (Peek *et al.*, 1998). Furthermore, recombination between vertically transmitted symbionts is limited to within the homogenous endosymbiont population (Stewart and Cavanaugh, 2006). Indeed, thioautotrophic symbionts vertically transmitted in bivalves have been observed to have higher 16S rDNA gene substitution rates compared to free-living Gammaproteobacteria, Betaproteobacteria, and other environmentally transmitted thioautotrophic marine symbionts, including those associated with lucinid clams (Peek *et al.*, 1998). The low evolution rate of horizontally transmitted symbionts can be attributed to purifying selection in larger free-living populations, in relation to intracellular host-associated populations (Peek *et al.*, 1998). Compared to vertical transmission, horizontal transmission could theoretically evolve based on other scenarios (Genkai-Kato and Yamamura, 1999). For instance, the cost of vertical transmission could be high, or symbionts could probably harm immature juveniles, as some studies with mycorrhiza fungi have suggested (Genkai-Kato and Yamamura, 1999). Horizontal transmission could also evolve under fluctuating environmental conditions where the host copes by sampling and domesticating bacteria best adapted to their current ecological niche (Won *et al.*, 2003; Ferdy and Godelle, 2005; Roeselers and Newton, 2012). Ecological specialization in

bacteria is, in turn, facilitated by horizontal gene transfer events (Papke and Gogarten, 2012).

Symbiont acquisition in lucinid clams

Unlike chemosymbiotic Solemyidae and Vesciomyidae clam families, where vertical or possibly mixed symbiont transmission have been observed (Stewart *et al.*, 2008; Roeselers and Newton, 2012), lucinid bivalves studied to date appear to capture their chemosynthetic symbionts from the environment. Chemosymbiont-specific primers tested on *Codakia orbicularis*, *Ctena orbiculata*, *Lucina pensylvanica*, *Lucinoma aequizonata*, *Parvilucina pectinella*, and *Phacoides pectinatus* amplified DNA targets from gill tissues but failed to amplify targets from reproductive tissues (Gros *et al.*, 1996b; Gros *et al.*, 1998; Gros *et al.*, 1999). Experiments on *C. orbicularis* juveniles reared in sterilized and unsterilized seagrass beds demonstrate that only those hosts grown in unsterilized seagrass beds were able to acquire chemosynthetic endosymbionts after larval metamorphosis (Gros *et al.*, 1996b; see also "Lucinid growth and reproduction" subsection above). Free-living forms of the *C. orbicularis* chemosymbionts, detected via fluorescence in situ hybridization (FISH), can infect juvenile clams (Gros *et al.*, 2003a). Juveniles of *C. orbicularis* can also acquire chemosymbionts from purified gill-symbiont sections of their own species, as well as other lucinid species hosting chemosymbionts with identical 16S rRNA gene sequences, including *Anodontia alba*, *Ctena orbiculata*, *Divaricella quadrisulcata*, *Linga pensylvanica*, and *Parvilucina pectinella* (Gros *et al.*, 2003b). This is probably because

undifferentiated naïve bacteriocytes precursors in juvenile clams do not discriminate against chemosymbiont strains (Gros *et al.*, 2003a; Brissac *et al.*, 2016).

On the other hand, experiments on another lucinid species, *Ctena orbiculata*, demonstrate that adult clams lose their chemosynthetic symbionts when starved for sulfide and could re-acquire endosymbionts from their natural habitat, but not in symbiont-free seawater with sulfide (Gros *et al.*, 2012). These findings suggest that horizontal symbiont acquisition happens throughout the lifespan of *C. orbiculata* (Gros *et al.*, 2012). A recent cross-infection experiment further shows that starved *C. orbiculata* adults could only re-acquire symbiont strains that they initially hosted (Brissac *et al.*, 2016). This observation led the authors to postulate that lucinid-symbiont evolutionary processes are antagonistic, as symbionts evolve to avoid being trapped and exploited within the clam gills, but that lucinids evolve to capture and farm their symbionts (Caro *et al.*, 2007; Brissac *et al.*, 2009; Brissac *et al.*, 2016). Despite the host's capability to reacquire chemosymbionts, experiments show that *C. orbiculata* and *Codakia orbicularis* adults do not release their chemosymbionts into the environment, which rules out the possibility of transgenerational symbiont inheritance via host-to-host symbiont transmission (Gros *et al.*, 2003b; Brissac *et al.*, 2009). Another histological study on *C. orbicularis* gill tissues reveals a heterogeneous endosymbiont population with predominantly large cell sizes and multiple copies of symbiont genomes, accompanied with rare occurrences of symbiont cell division observable under the electron microscope (Caro *et al.*, 2007). Based on the results, the authors speculate that *C. orbicularis* may inhibit endosymbiont multiplication in the gills (Caro *et al.*, 2007), akin to the lichen-algae (Ahmadjian, 1993), coral-*Symbiodinium*

(Woolridge, 2010), weevil-gammaproteobacterium (Login and Heddi, 2013), and legume-rhizobia (Udvardi and Poole, 2013) symbiotic systems.

Taken together, studies of symbiont transmission modes in lucinid bivalves have led to the hypothesis that the lucinid-bacteria chemosymbiosis may not be strictly mutualistic (Brissac *et al.*, 2009). Chemosynthetic endosymbionts in *C. orbicularis* and *Codakia orbiculata* are predicted to be trapped and exploited by their hosts, possibly in a form of "controlled parasitism," which would move towards an evolutionary dead end (Ahmadjian, 1993; Caro *et al.*, 2007; Brissac *et al.*, 2009; Woolridge, 2010; Brissac *et al.*, 2016). This agrees with the perspective that symbiont fitness may not always be increased in hosts (Garcia and Gerardo, 2014).

Phylogeny of chemosynthetic lucinid symbionts

All chemosynthetic lucinid endosymbionts described so far are thioautotrophic chemolithoautotrophs from a single bacterial class, Gammaproteobacteria (Distel *et al.*, 1988; Cavanaugh *et al.*, 2006). But, 16S rRNA phylogenies of chemosynthetic marine symbionts show at least nine separate lineages, which implies that the symbionts might have evolved independently multiple times from different free-living ancestors (Kleiner *et al.*, 2012; Dubilier *et al.*, 2008). To date, symbionts in this group are not yet cultured (Cavanaugh *et al.*, 2006; Dubilier *et al.*, 2008). Gammaproteobacterial lucinid symbionts, as well as other marine symbionts, have a vast array of metabolic capabilities, including thioautotrophy, mixotrophy, methanotrophy, carboxydotrophy, and hydrogenotrophy (Cavanaugh *et al.*, 2006; Woyke *et al.*, 2006; Petersen *et al.*, 2011; Kleiner *et al.*, 2012;

Dmytrenko *et al.*, 2014; Nakagawa *et al.*, 2014; Kleiner *et al.*, 2015). The 16S rRNA gene phylogenies of thioautotrophic marine symbionts are incongruent with phylogenies of their key metabolic genes, suggesting that horizontal gene transfer may have driven the convergence and divergence of their metabolic functions (Kleiner *et al.*, 2012). For instance, symbionts can utilize the Sox enzyme complex (lacking SoxCD), reverse dissimilatory sulfite reductase (Dsr) proteins, adenylylsulfate reductase (Apr), and sulfate adenylyltransferase (Sat) for the oxidation of sulfur, thiosulfate, sulfide, and sulfite (Ghosh and Dam, 2009; Kleiner *et al.*, 2012). Some symbionts, including lucinid symbionts, are also capable of using sulfide:quinone oxidoreductase (Sqr) for sulfide oxidation (Kleiner *et al.*, 2012; Petersen *et al.*, 2016). This stands in contrast to their free-living counterparts that have variable sulfur oxidation pathways, such as those involving SoxCD, which implies that convergent evolution of the sulfur oxidation pathways may have taken place in the marine thioautotrophic symbionts (Kleiner *et al.*, 2012). Most chemosynthetic marine symbionts perform carbon fixation with the Calvin-Benson-Bassham cycle, with variations in the form of RuBisCO enzyme and other CBB cycle enzymes, including sedoheptulose-1-7-bisphosphatase and fructose-1,6-bisphosphatase or pyrophosphate-dependent 6-phosphofructokinase (Robinson *et al.*, 1998; Kleiner *et al.*, 2012). Besides the Calvin-Benson-Bassham cycle, symbionts from the tubeworm *Riftia pachyptila* are also capable of carbon fixation via the reductive tricarboxylic acid cycle (Markert *et al.*, 2007; Markert *et al.*, 2011; Gardebrecht *et al.*, 2012). Although some thioautotrophic symbionts are obligate autotrophs, others exhibit potential heterotrophy or mixotrophy (Woyke *et al.*, 2006; Kleiner *et al.*, 2012; Dmytrenko *et al.*, 2014; Petersen *et al.*, 2016).

At the 16S rRNA gene sequence level, chemosynthetic lucinid endosymbionts form a paraphyletic group, where some are most closely related to sulfur-oxidizing endosymbionts in *Solemya* and *Thyasira* clams, while others are closer to the siboglinid tube worms such as *Ridgeia piscesae* and *R. pachyptila* (Cavanaugh *et al.*, 2006). Based on their 16S rRNA gene sequences, chemosynthetic lucinid endosymbionts have been previously classified into three clades, with each clade corresponding to a single species (Brissac *et al.*, 2011). The largest clade, clade A, consists of chemosynthetic symbionts from predominantly seagrass-dwelling lucinid bivalves, and is further subdivided into two subgroups (Brissac *et al.*, 2011). Clade A symbionts share >97% to 100% 16S rRNA sequence identity with each other (Durand and Gros, 1996; Durand *et al.*, 1996; Gros *et al.*, 2003; Brissac *et al.*, 2011; Brissac *et al.*, 2016). On the other hand, chemosynthetic symbionts from clades B and C are from the mangrove-dwelling lucinid species *Anodontia* spp. and *Phacoides pectinatus*, respectively (Brissac *et al.*, 2011). The low to zero variability in 16S rRNA gene sequences, especially among clade A symbionts, is surprising because lucinid bivalves are geographically diverse and acquire their chemosynthetic symbionts horizontally (see also "Habitats of lucinid clams" and "Symbiont acquisition in lucinid clams" subsections, above). Therefore, chemosynthetic lucinid symbionts should presumably possess higher heterogeneity than vertically transmitted symbionts (see also "An overview of symbiont transmission modes" subsection, above; Brissac *et al.*, 2011). To explain this phenomenon, Brissac *et al.* (2011) suggest that, rather than co-evolution, the association between lucinid clams and their symbionts is opportunistic and dictated by the type of bacterial species in the environment. Analogously, low endosymbiont genetic

variability has also been observed between *Crysomallon squamiferum* snail individuals, which is attributed to strict symbiont selection by the host to overcome genetic drift in the intracellular, horizontally-acquired endosymbiont population that has a smaller genome size than that of the free-living populations (Nakagawa *et al.*, 2014).

Nevertheless, molecular symbiont phylogeny based on the slow-evolving 16S rRNA gene does not provide sufficient strain-level resolution of chemosynthetic lucinid symbionts (Brissac *et al.*, 2016). Recent comparative analysis of six lucinid species in the Caribbean harboring clade A symbionts, using the 16S-23S rRNA gene internal transcribed spacer region, adenylylsulfate reductase alpha subunit (*aprA*), type I RuBisCO large chain (*cbbL*), DNA polymerase III subunit alpha (*dnaE*), and DNA gyrase subunit B (*gyrB*) marker genes, reveals unprecedented strain-level diversity that is non-randomly distributed to a certain extent by host geographic location (Brissac *et al.*, 2016). Because there is great diversity in lucinid morphologies and habitats, it is conceivable that different symbiotic strains and species exist, and that they can utilize different metabolic pathways (Taylor and Glover, 2000).

Lucinid gill microbiome diversity

Beyond the chemosynthetic gill endosymbionts, bacterial taxonomic diversity in the gills of lucinid bivalves is understudied, although dual or multiple symbionts have been reported from other chemosynthetic marine organisms. For instance, dual symbiosis in Mytilidae hydrothermal vent and cold seep mussels with thioautotrophic and methanotrophic symbionts was determined through 16S rRNA gene sequence analysis,

FISH, and/or transmission electron microscopy (Distel *et al.*, 1995; Duperron *et al.*, 2005; Duperron *et al.*, 2006). The dual symbiosis enables these mussels to occupy habitats that have high levels of sulfide and/or methane (Duperron *et al.*, 2006). Besides the thioautotrophic and methanotrophic symbionts, intracellular Oceanospirillales, Rickettsia- and Chlamydia-like parasitic bacteria have also been described from deep-sea bathymodiolin mussels (Zielinski *et al.*, 2009). Particularly, microscopy analyses of the Oceanospirillales species reveals the bacterium is a parasite that multiplies in nuclei of non-bacteriocytes, which causes nuclear lysis and bacteria release (Zielinski *et al.*, 2009). Using similar approaches, endosymbiotic communities in the gutless marine worms *Olavius* spp. consist of sulfide-oxidizers, sulfate-reducers, and spirochetes (Dubilier *et al.*, 2001; Blazejak *et al.*, 2005; Woyke *et al.*, 2006; Ruehland *et al.*, 2008). In these worms, the sulfate-reducers and sulfide-oxidizers engage in a mutualistic relationship through the recycling of sulfur compounds (Dubilier *et al.*, 2001).

Similar 16S rRNA gene analyses and/or microscopy studies on lucinid clams show that the gill microbiomes consist of other species in addition to the chemosynthetic symbionts. In the gills of *Lucinoma aff. kazani*, 16S rRNA gene phylotypes of the thioautotrophic symbiont and another *Spirochaeta*-like species related to free-living *Spirochaeta coccoides* have been documented (Duperron *et al.*, 2007). Extracellular spirochete-like bacteria (8–10 μm length and 0.2–0.3 μm width), intracellular chemosynthetic symbionts (3–5 μm length and 0.5–1.0 μm width), and another intracellular rod-shaped bacterium (1 μm length and 0.4–0.5 μm width) have also been visualized in the gills of *Anodontia ovum* (Ball *et al.*, 2009). In the gills of *Loripes*

orbiculatus, two 16S rRNA gene phylotypes classifiable to the chemosynthetic symbiont and another gammaproteobacterial species grouped with non-symbiotic species like those in marine sediments have also been identified (Espinosa *et al.*, 2013). Notwithstanding marker gene sequence and microscopy evidence, the prevalence, functions, and details about the symbiont and/or host association of these non-chemosynthetic bacteria remain unknown.

Dissertation objectives

Research efforts on the lucinid-bacteria chemosymbiosis reviewed so far have centered mainly on paleontology, host habitat, host reproduction, host nutrition, gill morphology, chemosymbiont functions, chemosymbiont transmission, and chemosymbiont diversity. Despite extensive studies on this symbiotic system, several gaps remain in the literature. First, although sequencing studies of the chemosynthetic symbionts have focused on the 16S rRNA gene (see "Lucinid gill microbiome diversity" subsection, above), recent investigations are just beginning to use additional marker genes to resolve symbiont strain diversity (Brissac *et al.*, 2016), as well as -omics approaches to identify the range of functions possible in lucinid chemosymbionts (König *et al.*, 2016; Petersen *et al.*, 2016). To date, the genomes of only two clade A chemosynthetic lucinid symbiont species in *Codakia orbicularis* (König *et al.*, 2016) and *Loripes orbiculatus* (referred to as *Loripes lucinalis* in Petersen *et al.*, 2016) have been sequenced (Petersen *et al.*, 2016), and only the transcriptome of the *L. orbiculatus* chemosymbiont has been sequenced (Petersen *et al.*, 2016). This results in an insufficient understanding of the genetic and metabolic

repertoire of chemosynthetic symbionts, especially those belonging to clades B and C, as well as how the symbionts vary within and between species, among the host species, and within host populations. Second, although studies on host nutrition combined with geochemistry, gill morphology, and symbiont acquisition provide useful insights into host behavior (Rossi *et al.*, 2013; see also "Lucinid growth and reproduction" subsection above), lucinid bivalve genomes are not yet available, which is hampering efforts to examine host functions at the genetic level. Third, in spite of previous findings that suggest taxa-diverse species may coexist with the chemosynthetic symbionts in lucinid clam gills, similar to other chemosymbiotic marine organisms (see also "Lucinid gill microbiome diversity" subsection, above), high-resolution -omics approaches have not been used to comprehensively re-examine gill microbiome diversity in lucinid bivalves. Lastly, although previous studies also highlight important roles of lucinid habitats in facilitating symbiotic functions and structuring symbiotic diversity (see "Lucinid Habitats" and "Lucinid-bacteria-environment interactions" subsections, above; Brissac *et al.*, 2016), there is a dearth of integrative studies that investigate the influences of environmental parameters on microbiome diversity, microbial functional potential, and host-microbiome gene expression. Overall, these literature gaps lead to a poor understanding of the taxonomic, genetic, and functional complexity of lucinid gill microbiomes, the range of microbiome (including the chemosynthetic symbiont) and host functions in lucinid gills, as well as potential environmental drivers shaping microbiome diversity and host-microbiome gene expression.

This dissertation aims to fill in some of these knowledge gaps by comprehensively examining the taxonomic, genetic, and functional diversity in the gill microbiomes of a mangrove-dwelling lucinid species, *Phacoides pectinatus*, and two tropical seagrass lucinid species, *Ctena orbiculata* and *Stewartia floridana*, using 16S rRNA gene sequencing, qPCR, metagenomic sequencing, and metatranscriptomic sequencing. Besides microbiome characterization, this dissertation also aims to identify and quantify host-related transcripts in the gills of lucinid bivalves. Finally, this dissertation seeks to explore the effects of environmental drivers, such as seagrass coverage, sulfide concentrations, and oxygen concentrations, on lucinid host and symbiont gene expression in seagrass-associated *C. orbiculata* and *S. floridana*.

Hypotheses

Hypothesis 1a – Gill microbiomes of *P. pectinatus*, *C. orbiculata*, and *S. floridana* are taxonomically diverse.

Previous Sanger and 454 pyrosequencing of 16S rRNA gene sequences in the gills of *P. pectinatus* by our collaborators reveal unprecedented symbiont taxonomic diversity within the lucinid species (Green- García, 2008; Doty, 2015). Nearly full-length 16S rRNA gene sequence analysis of *P. pectinatus* collected from shallow sea grass (*Thalassia testudinum*) beds in the Mouth of Pigeon Creek at the Bahamas has identified a species most closely related to a methane-utilizing alphaproteobacterial *Methylobacterium* spp. clone (Green-García, 2008). In the gills and feet of *P. pectinatus* collected in 2014 at

Wildcat Cove, Florida, analysis of the V1-V3 region of the 16S rRNA gene reveals the presence of bacteria belonging to the genera *Sedimenticola*, *Kistimonas*, *Methylomarinum*, *Spirochaeta* spp., as well as unclassified Rickettsiales and Lentisphaerae (Doty, 2015). *Phacoides pectinatus* gills with higher abundances of Rickettsiales were isolated from areas with the lowest clam density, whereas gills with higher *Methylomarinum* abundances were collected from deeper sediments (Doty, 2015). Using 454 pyrosequencing, our collaborators also analyzed the V1-V3 region of the 16S rRNA gene in gills of *S. floridana* collected at Bokeelia Pier, Florida, in 2014 (Goemann, 2015). Sequence analysis indicates the presence of other taxa, besides the dominant thioautotrophic symbiont, in *S. floridana* gills (Goemann, 2015). These preliminary results from *P. pectinatus* and *S. floridana* led to the hypothesis that gill microbiomes in these species are more taxonomically and functionally diverse than previously thought. This hypothesis is also extended to *C. orbiculata* collected from Sugarloaf Key, Florida, collected in 2016. A combination of qPCR, 16S rRNA gene, metagenomic, and metatranscriptomic analyses will be used to test the hypotheses. This is an important area of study because high-throughput 16S rRNA gene sequencing has not been used to investigate gill microbiome diversity in lucinid clams systematically.

Hypothesis 1b – Taxonomically diverse gill microbes in *P. pectinatus*, *C. orbiculata*, and *S. floridana* confer novel metabolic capabilities to the gill microbiomes.

This hypothesis predicts that lucinid gill microbiome members, if taxonomically diverse, are metabolically distinct from each other. For instance, the identification of *Methylobacterium*-like and *Methylomarinum*-like OTUs in *P. pectinatus*, if supported, would confer novel methylotrophic functions to the gill microbiomes (Green-García, 2008; Doty, 2015). Metagenomic and metatranscriptomic approaches will be used to survey the metabolic profiles of lucinid gill microbiome species. PCR and/or qPCR will be performed, if necessary, to validate, the presence and activity of metabolic genes of interest in the lucinid gill microbiomes.

Hypothesis 2 – Chemosynthetic symbionts in *P. pectinatus*, *C. orbiculata*, and *S. floridana* encode and express novel metabolic genes not yet discovered in chemosynthetic lucinid symbionts.

Because genomic data of chemosynthetic lucinid symbionts is currently limited to clade A symbionts from *Codakia orbicularis* and *Loripes orbiculatus* (König *et al.*, 2016; Petersen *et al.*, 2016), this hypothesis predicts that metabolic functions in chemosynthetic lucinid symbionts are under-sampled and that metagenomic and metatranscriptomic analyses would uncover previously undescribed genes and pathways corresponding to host and/or habitat differences in the clade C symbiont of mangrove-dwelling *P. pectinatus*, as well as clade A symbionts of seagrass-dwelling *Ctena orbiculata* and *S. floridana*. As with hypothesis 1b, a combination of metagenomic, metatranscriptomic, and PCR/qPCR approaches will be used to test the hypothesis.

Hypothesis 3 – Environmental factors affect gene expression of lucinid clams and their thioautotrophic symbionts.

To date, environmental effects on symbiont and host gene expression in lucinid species have not been studied, although they have been demonstrated to influence host and gene expression in other marine symbiotic systems (Scott *et al.*, 2004; Girguis and Childress, 2006; Carney *et al.*, 2007; Boutet *et al.*, 2011; Duperron *et al.*, 2011; Robidart *et al.*, 2011; Beinart *et al.*, 2012; Gardebrecht *et al.*, 2012; Kleiner *et al.*, 2015). Preliminary analyses performed by our collaborators on *S. floridana* specimens collected at Bokeelia Pier, Florida, in 2014, and their seagrass habitat reveal large areas of stable habitat over the past seven years (Long, 2016). Considering that the maximum age of the live specimens collected were around six years old, the results of geospatial analysis suggest that *S. floridana* clams at Bokeelia Pier experience a consistent habitat throughout their life span (Long, 2016). The population densities of *S. floridana* are also higher in areas with high seagrass coverage than in bare sand patches (Long, 2016). Furthermore, *S. floridana* collected from seagrass areas and sand patches exhibit significant differences in terms of shell morphology, which suggests that there may be potential effects of seagrass coverage on the clam and chemosynthetic symbiont functions (Long, 2016). Based on research on other marine symbiosis systems and our collaborators' preliminary analyses, I hypothesize that functional differences between symbionts from the same taxonomic group (e.g., gammaproteobacterial thioautotrophic symbionts) across lucinid species can also be influenced by host mechanisms, as well as environmental factors, such as vegetation cover

and substrate (e.g., oxygen and sulfide) availability. This hypothesis will only be tested on the lucinids *C. orbiculata* and *S. floridana*, in relation to potential lucinid-bacteria-seagrass interactions (see also "Lucinid-bacteria-environment interactions" subsection, above). Because the lucinid samples were collected from the field where environmental conditions are dynamic and cannot be controlled, for each collection site, I will analyze gill metatranscriptomes of samples collected from at least two quadrats that contrast in at least one environmental parameter. The null hypothesis is that environmental differences will not affect symbiont and host gene expression. If the null hypothesis is supported, then I will not observe any significant differences in symbiont and host gene expression in samples collected from the contrasting quadrats. If the null hypothesis is rejected, then I will detect significant differences in the expression levels of at least one gene belonging to the symbionts and/or their hosts.

CHAPTER TWO

TAXONOMIC AND FUNCTIONAL HETEROGENEITY OF THE GILL MICROBIOME IN A SYMBIOTIC COASTAL MANGROVE LUCINID SPECIES

Citation

Lim SJ, Davis BG, Gill DE, Walton JL, Nachman E, Engel AS, Anderson LC, Campbell BJ (2018). Taxonomic and functional heterogeneity of the gill microbiome in a symbiotic coastal mangrove lucinid species. ISME J, doi: 10.1038/s41396-018-0318-3.

Authors

Shen Jean Lim[1], Brenton G. Davis[1,2], Danielle E. Gill[1,3], Jillian Walton[1], Erika Nachman[1], Annette Summers Engel[4], Laurie C. Anderson[5], and Barbara J. Campbell[1]

[1]Department of Biological Sciences, Clemson University, Clemson, SC 29634-0001

[2]Current address: College of Medicine, Medical University of South Carolina, Charleston, SC 29425-8900

[3]Current address: Instituto de Medicina Tropical São Paulo, Universidade de São Paulo, São Paulo 05403-000

[4]Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN 37996-1410

[5]Department of Geology and Geological Engineering, South Dakota School of Mines and Technology, Rapid City, SD 57701-3901

Authors' contributions

A.S.E., B.J.C., and L.C.A. secured the funding for this study, supervised sample collection, and put in research efforts; S.J.L., B.J.C., A.S.E., L.C.A. collected the samples used in the study; S.J.L. and B.J.C. conceived the experiments; S.J.L. performed most of the experiments, software implementation, data analyses, and wrote this chapter. B.G.D. performed qPCR analyses on the thioautotrophic symbiont and *Kistimonas*-like species; D.E.G. and J.W. performed qPCR and PCR analyses on the *Spirochaeta*-like species; J.W. performed transcriptomic analyses on the *Spirochaeta*-like species; E.N. performed PCR analyses on the *Kistimonas*-like species; S.J.L. maintains the NCBI sequence data and L.C.A. curates the metadata and maintains specimens of dissected tissues and valves. B.J.C, A.S.E., and L.C.A. reviewed and edited this chapter.

Introduction

Chemosymbiosis is widespread in marine habitats, where endo- or epi-symbiotic chemolithoautotrophs use inorganic chemical energy for the synthesis of organic compounds that benefit their hosts (Dubilier *et al.*, 2008; Taylor and Glover, 2010). One of the most ancient examples of marine chemosymbiosis is found in the bivalve family Lucinidae (Taylor and Glover, 2000), which has a fossil record arguably dating back to the Silurian period (Liljedahl, 1992). Despite being capable of suspension feeding, all living lucinids studied to date fulfill a considerable proportion of their nutritional needs through obligate chemosymbiotic associations with gammaproteobacterial endosymbionts occupying bacteriocytes in their gills (Taylor and Glover, 2000). Lucinid species examined

so far acquire their thioautotrophic endosymbionts from free-living environmental bacterial populations (Gros *et al.*, 1996b; Gros *et al.*, 1998; Gros *et al.*, 1999; Gros *et al.*, 2003; Brissac *et al.*, 2009). Enzymatic assays, stable isotope analyses, and clone-based amplicon sequencing methods demonstrate that lucinid endosymbionts mainly use energy derived from the oxidation of reduced sulfur compounds to fix inorganic carbon for their hosts (Cavanaugh *et al.*, 2006). Other reported functions of lucinid endosymbionts included mixotrophy, denitrification, assimilation of nitrogenous compounds, and diazotrophy (Fisher and Hand, 1984; Hentschel and Felbeck, 1995; Petersen *et al.*, 2016; König *et al.*, 2016).

Because of the widespread distribution of lucinids in marine habitats, ranges in host and endosymbiont phylogenetic diversity, as well as the possibility that lucinids may harbor non-thioautotrophic symbionts (Ball *et al.*, 2009; Duperron *et al.*, 2012; Pales Espinosa *et al.*, 2013), the lucinid-bacteria chemosymbiotic system has the potential to address fundamental cellular to ecological questions about host-symbiont interactions, cues, and communication across individual hosts, among species, and within populations. However, there is still relatively poor understanding of lucinid and gill microbiome diversity and metabolic functions. For instance, although 16S rRNA gene sequences of thioautotrophic lucinid endosymbionts form a paraphyletic group consisting of three distinct clades (Brissac *et al.*, 2011; Cavanaugh *et al.*, 2006), only the genomes, transcriptomes, and proteomes of two lucinid endosymbiont species from clade A have been sequenced (Petersen *et al.*, 2016; König *et al.*, 2016). Clade A symbionts are associated predominantly with diverse seagrass-dwelling lucinids, but symbiont clades B

and C are from predominately mangrove-dwelling *Anodontia* spp. and *Phacoides pectinatus*, respectively (Brissac *et al.*, 2011). Almost no diversity or functional diversity study has centered on either of these bacterial clades.

To begin to fill these gaps, our study characterizes the metabolic repertoire of the host and gill-associated thioautotrophic bacterial endosymbiont from *Phacoides pectinatus* Gmelin 1791 (syn = *Tellina pectinata* Gmelin 1791, *Lucina pectinata* (Gmelin 1791), *Anodontia pectinatus* (Gmelin 1791), *Lucina jamaicensis* Lamarck 1801, *Lucina funiculata* Reeve 1850). Possibly the only extant species of its genus, *P. pectinatus* possesses morphological features distinct from other lucinid bivalves, such as high levels of three types of hemoglobin in gill pigment granules, sulfur bodies, and large lysosomes (Read, 1965; Liberge *et al.*, 2001). Molecular phylogeny studies place *P. pectinatus* as a deeply-branching genus within the Lucinidae (Williams *et al.*, 2004) and the thioautotrophic endosymbiont distant from seagrass- or other mangrove-associated lucinid endosymbionts (Brissac *et al.*, 2011; Durand *et al.*, 1996; Green-García, 2008). This lucinid inhabits organic-rich seagrass and mangrove sediments (Frenkiel *et al.*, 1997), and has a widespread tropical geographic distribution that ranges from the Caribbean Sea and Gulf of Mexico, to the Atlantic Ocean seaboard of South America to Brazil (Christo *et al.*, 2016). The unusual morphological features, phylogeny, and habitat distribution of *P. pectinatus* and its distinct thioautotrophic endosymbiont belonging to clade C have led to the hypothesis that symbiont metabolic pathways in this species are different than in other lucinid endosymbionts (Gros *et al.*, 1998). To test this hypothesis, we assessed gill microbiome diversity within *P. pectinatus* using 16S rRNA gene sequencing, quantitative

PCR (qPCR), metagenomic sequencing, and metatranscriptomic sequencing, and compared the expression profiles from *P. pectinatus* and its gill microbiome species to previously sequenced seagrass-associated lucinid endosymbiont species from clade A, including *Ca.* Thiodiazotropha endoloripes within *Loripes orbiculatus* (Petersen *et al.*, 2016) and *Ca*. Thiodiazotropha endolucinida within *Codakia orbicularis* (König *et al.*, 2016).

Materials and methods

Sample collection

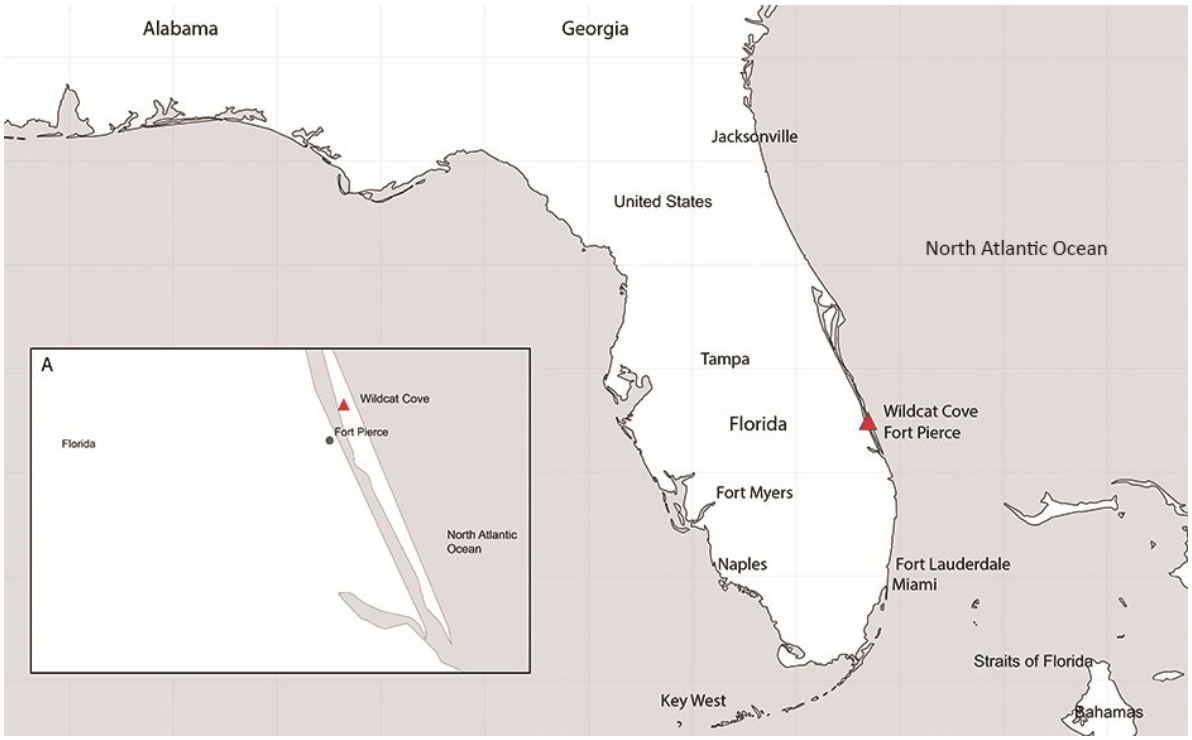


Figure 2.1. Map showing location of the sampling site, with a close-up view of Wildcat Cove, Florida, USA (Insert A).

Phacoides pectinatus populations at Wildcat Cove, St. Lucie County, Florida, USA (Figure 2.1), as well as their ecology, sediment geochemistry, and microbiology have previously been investigated (Green-García, 2008; Doty, 2015). After lucinid density at Wildcat Cove was established to be more than one *Phacoides pectinatus* specimen per shovel-full of sediment, which was within 3 m of the mangrove-lined shoreline, a sampling area was sectioned off into quadrats, ranging from 0.5 m^2 to 1 m^2 over, in general, a 100 m^2 area. During the 2014 sampling, sediment porewater was obtained from six quadrats (Doty, 2015) by low-flow fluid sampling using stainless steel piezometers installed near to where specimens were recovered, based on previously descried methods (Green-García and Engel, 2012). Standard electrode methods were used to measure dissolved oxygen content from the porewater, as well as pH, temperature, conductivity, and to collect water samples for dissolved ion and total organic carbon concentrations (Doty, 2015). Dissolved sulfide concentrations were obtained colorimetrically using CHEMetrics (Calverton, VA, USA) chemistry and a field spectrophotometer (Green-García and Engel, 2012; Doty, 2015).

For this study, research excursions were completed in February 2011, June 2013, July 2014, and November 2017, and live specimens were sieved from sediments hand-dug to 30 cm depth, approximately 3 m from the shoreline of *Rhizophora mangle* (red mangrove). Specimens were temporarily stored in Whirl-Pak® Bags (Nasco, Fort Atkinson, WI, USA) filled with surface water from the habitat and maintained at ambient temperature before dissection. During dissection, gill and foot tissues were separated from other body tissues. Tissues used for 16S rRNA gene sequencing and metagenomics were dissected

within the same day of collection and fixed in 100% molecular grade ethanol. Tissues used for metatranscriptomics were dissected within 30 minutes of collection and fixed in RNAlater. Tissues used for microscopy were fixed in 2% paraformaldehyde (pH 7) made with artificial sea water prepared using Difco™ Marine Broth 2216 formula (Becton Dickinson and Company, Franklin Lakes, NJ, USA) for 3 hours at 4°C prior to washing, sucrose infiltration, storage, hematoxylin-eosin (H&E) staining, and fluorescence in situ hybridization (FISH) procedures. H&E stained sections were visualized using Leica's DM750 microscope attached to a Leica ICC50HD camera and the LAS EZ V2.1.0 software (Buffalo Grove, IL, USA).

Fluorescence in situ hybridization (FISH)

Following paraformaldehyde fixation, gill tissues were washed 3x for five minutes each in artificial sea water (ASW) and stored in the same medium overnight before infiltration with 10% and 25% sucrose in ASW. Tissues were stored at 4°C prior to cryosectioning. Hematoxylin-eosin (H&E) staining and FISH were performed on 5 µm cryosections from tissues cut with Thermo Fisher Scientific's (Pittsburgh, PA, USA) Microm™ HM550 Cryostat on VistaVision™ HistoBond® Adhesive Slides (VWR, Radnor, PA, USA). Prior to FISH, cryosections were soaked in diethyl pyrocarbonate (DEPC)-treated phosphate-buffered saline (PBS) solution for five minutes and air dried for an hour. Hybridization, washing, counter-staining, and mounting steps were performed in accordance to the Standard FISH protocol (https://www.arb-silva.de/fish-probes/fish-protocols/) available on Silva's web server (Quast *et al.*, 2013), except that 4',6-diamidino-

2-phenylindole (DAPI) counter-staining was extended to 12 minutes, followed by three washes at five minutes each with DEPC-treated PBS solution.

A probe named SED642 (5'-ACCATACTCTAGCCTGCCAG-3') was designed to hybridize to the *P. pectinatus* endosymbiont, *Ca*. Sedimenticola endophacoides, based on the alignment of full-length 16S rRNA gene sequences extracted from the species' MAGs with the BangT-642 probe used for the *Bathymodiolus* sp. mussel symbiont (5′-CCTATACTCTAGCTTGCCAG-3′; Duperron *et al.*, 2005) in ClustalW (Thompson *et al.*, 1994) implemented in BioEdit 7.2.5 (Hall, 1999). The specificity of probe SED642 was evaluated using the NCBI's Basic Local Alignment Search Tool (BLAST) web tool (Altschul *et al.*, 1990) against the 16S ribosomal RNA sequence database (NCBI Resource Coordinators, 2016). Probe SED642 shared 100% sequence identity with 16S rRNA gene sequences of strains from the genera *Salinispirillum*, *Methylophaga*, *Marinomonas*, *Methylosphaera*, and *Pseudomonas*, and one or more mismatches to a range of other strains predominately within the class Gammaproteobacteria, but also to strains from the Actinobacteria and Firmicutes (Table 2.1). As genera matching probe SED642 have not been previously associated with symbiotic associations in bivalves, the likelihood of false positive signals from these organisms on the *P. pectinatus* gill samples was considered low. Probe SED642 was labelled with Cy5 (Integrated DNA Technologies, Skokie, IL, USA) at the 5' end. Probes EUB338 (5'-GCTGCCTCCCGTAGGAGT-3'; Amann *et al.*, 1990) for the general detection of bacteria and its reverse complement NON338 (5'-ACTCCTACGGGAGGCAGC-3'; Wallner *et al.*, 1993) were both labelled with Cy3 (Integrated DNA Technologies, Skokie, IL, USA) at the 5' end and used as positive and

negative controls, respectively. A range of formamide concentrations between 10% to 35% were tested and signals for all probes were found to be optimal at 30% formamide concentration. Hybridized sections were imaged with Leica's SPE confocal microscope at the Clemson Light Imaging Facility (Clemson, SC, USA) and the Leica Application Suite (LAS) X software (Buffalo Grove, IL, USA).

16S rRNA gene, metagenomic, and metatranscriptomic sequencing

Total nucleic acids were extracted from partial gill and foot tissues using Qiagen's (Valencia, CA, USA) DNeasy Blood and Tissue kit (2011 and 2013 sample collection) or Allprep DNA/RNA Mini Kit (2014 samples) after mechanical homogenization of the tissues with a motorized pestle and mortar (Argos Technologies, Elgin, IL, USA) or tissue grinder (Wheaton, Millville, NJ, USA). For further lysis, the sample was passed through a 21-gauge (0.8mm) needle attached to a 3 mL syringe (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) at least ten times and incubated at 60°C for at least ten minutes. Extracted nucleic acid concentrations were quantified fluorometrically with Qubit™ dsDNA HS and RNA assays (Life Technologies, Austin, TX, USA). From the 2014 collection, 16S rRNA gene libraries of DNA extracted from 25 *P. pectinatus* gills, cDNA from the gills of four of these individuals, and DNA from the feet of three individuals were sequenced by the Duke Center for Genomic and Computational Biology (Durham, NC, USA). From the 2017 collection, libraries of DNA and cDNA extracted from three gill samples were sequenced at Clemson University, SC, USA. All libraries were sequenced on Illumina Inc's (San Diego, CA, USA) MiSeq 2x250bp platform.

cDNA was synthesized from DNase-treated (Ambion® Turbo DNA-free™ DNase kit, Life Technologies) RNA with the High-Capacity cDNA Reverse Transcription Kit (Applied iosystems, Foster City, CA, USA). Successful DNase treatment and cDNA synthesis was confirmed with PCR amplification of the V9 region of bacterial 16S rRNA genes (see next section). DNA and cDNA from each sample was diluted to 1 ng/μL with nuclease-free water and amplified with previously developed dual indexes (Kozich *et al.*, 2013). Amplicons were normalized with the SequalPrep™ Normalization Plate Kit, 96-well (Invitrogen, Carlsbad, CA, USA), quantitated with the Qubit™ dsDNA HS assay (Life Technologies, Austin, TX, USA) and pooled for sequencing.

Table 2.1. Sequence comparisons of the *Ca.* Sedimenticola endophacoides SED642 fluorescence in situ hybridization (FISH) probe designed in this study with 16S rRNA gene sequences of other related bacterial species obtained from NCBI's 16S ribosomal RNA sequence database (NCBI Resource Coordinators, 2016).

Probe and 16S rRNA gene target	Sequence (5' → 3')
SED642 probe for *Ca*. Sedimenticola endophacoides (this study)	ACCATACTCTAGCCTGCCAG
BangT-642 probe for thiotrophic symbiont of *Bathymodiolus* sp. (Duperron *et al.*, 2005)	CCT----------T------
Salinispirillum marinum GCWy1 (NR_134169)[a]	--------------------
Methylophaga nitratireducenticrescens JAM1 (NR_074321)	--------------------
Marinomonas arenicola KMM 3893 (NR_112826)	--------------------
Marinomonas rhizomae IVIA-Po-145 (NR_116233)	--------------------
Marinomonas arctica 328 (NR_043882)	--------------------
Methylophaga alcalica M39 (NR_028824)	--------------------
Pseudomonas amygdali AL1 (NR_036999)	--------------------
Methylosphaera hansonii AM6 (NR_026033)	--------------------
Burkholderia singularis LMG 28154 (NR_152632), other *Burkolderia* strains[b]	X--------------------
Paraburkholderia caffeinilytica strain CF1 (NR_152088) and other *Paraburkholderia* strains	X----------C-----
Colwellia meonggei MA1-3 16S (NR_133732), and other *Umboniibacter* and *Solobacterium* strains	XXX----------------
Oceanospirillum beijerinckii subsp. *pelagicum* IFO 13612 (NR_112017), and other *Oceanospirillum*, *Oceanobacter*, and *Vibrio* strains	XXXX---------------
Yimella radicis py1292 (NR_152030), and other *Yimella*, *Calidifontibacter*, and *Neisseria* strains	----------XX
Pseudomonas cerasi 58 (NR_146827) and other *Pseudomonas, Methyloparacoccus* strains	----------U------
Methylocaldum marinum S8 (NR_126189) and other *Methylocaldum*, *Marinobacter*, and *Endothiovibrio* strains	----C---------------
Thiohalomonas denitrificans HLD 2 (NR_044097)	----A---------------
Methylohalobius crimeensis 10Ki (NR_042198)	----G---------------
Thioalkalispira microaerophila ALEN 1 (NR_025239), misc. *Pseudomonas* strains	----------U-------

[a] -, identical to probe sequence
[b] X, no base pair reported

Twelve Illumina-compatible gill metagenomic libraries and one foot metagenomic library were prepared using the Nextera DNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA) on 50 ng of DNA per sample (2011 and 2013 collection) by Molecular Research LP (Shallowater, TX, USA) and sequenced on Illumina's MiSeq 2x150bp (2011 collection), 2x250 bp platforms (2013 collection) and HiSeq 2500 2x100bp (2014 specimen) platforms. For deep sequencing, one Illumina-compatible gill metagenomic library from the 2014 collection was prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® on DNA fragmented with NEBNext® dsDNA Fragmentase (New England Biolabs, Ipswich, MA, USA; 2014 collection). For this library, insert size determination with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and outsourcing of Illumina HiSeq 2500 sequencing were performed by Clemson University Genomics Institute (CUGI; Clemson, SC, USA). To generate long reads, metagenomic libraries were prepared from another 2014 gill specimen and two 2017 gill specimens using Nanopore's Rapid Sequencing Kit (Oxford Nanopore Technologies, Kidlington, Oxfordshire, UK) and sequenced on a MinIon flowcell (R9.4 nanopores) with a MinIon Mk1B sequencer.

Three gill samples collected in 2014 within a $1m^2$ quadrat were used for metatranscriptomic sequencing on Illumina's HiSeq 4000 2x150bp platform. RNAs extracted from these samples were treated with the Ambion® Turbo DNA-free™ DNase kit (Life Technologies). To check for successful DNase treatment, the V9 region of the 16S rRNA gene in DNAse-treated (Ambion® Turbo DNA-free™ DNase kit, Life Technologies). RNA samples were amplified with previously described universal bacterial

primers, 1369F and 1492R (Suzuki *et al.*, 2000; Table 2.2). PCR reactions were performed in a 10 µl volume consisting of 0.25 µM of each primer and 1x BIO-X-ACT™ Short Mix (Bioline, Taunton, MA, USA). PCR amplification was carried out in the C1000 Touch™ Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA) under the following conditions: Initial denaturation at 95°C for 3 minutes, 29 cycles of denaturation at 95°C for 15 seconds, annealing at 53°C for 30 seconds, extension at 72°C for 30 seconds, followed by elongation at 72°C for 5 minutes and cooling at 12°C. Amplicons were run on a 1% (wt/vol) agarose gel, which was stained with ethidium bromide, de-stained with deionized water and visualized under UV with a FOTO/Analyst® FX System (Fotodyne Inc., Hartland, WI, USA).

DNA-free RNAs were purified with the RNeasy MinElute Cleanup Kit (Qiagen). Illumina-compatible cDNA libraries were made from purified RNAs using Epicentre's (Madison, WI, USA) Ribo-Zero rRNA removal kit (bacteria) and ScriptSeq™ v2 RNA-Seq Library Preparation Kit, following the manufacturer's low input protocol. The final concentration of each library was quantified with the Qubit® dsDNA HS assay (Life Technologies) and the average library insert size was determined with the Experion Automated Electrophoresis Station (Bio-Rad Laboratories, Hercules, CA, USA; 2011 and 2013 collections) and the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA; 2014 and 2017 collections). Metatranscriptomic libraries were pooled and sequenced by the Duke Center for Genomic and Computational Biology.

To check for DNA contamination in sequenced metatranscriptomic libraries, trimmed reads were mapped to a representative, high-quality *Ca.* Sedimenticola

endophacoides MAG (N3+P5) with high completeness value and low contamination and strain heterogeneity values (Table 2.3) using Bowtie2 v2.2.7's (Langmead and Salzberg, 2012) no-mixed, no-discordant, end-to-end, -k 200 and -gbar 1000 options. The read-to-MAG mapping was inspected in SeqMonk v1.42.0 (Babraham Bioinformatics, 2007) for consistent alignments across gene boundaries and directionality bias indicative of DNA contamination. No DNA contamination was observed in the metatranscriptomic libraries.

Table 2.2. List of PCR and qPCR primers used in this chapter.

Primer	**Sequence (5'->3')**	**Annealing temperature**	**Reference**
Universal 16S rRNA gene primer 1369F	CGGTGAATACGTTCYCGG	53°C	Suzuki *et al.*, 2000
Universal 16S rRNA gene primer 1492R	GGWTACCTTGTTACGACTT	53°C	Suzuki *et al.*, 2000
Universal M13 forward (-21) primer	GTAAAACGACGGCCAG	55°C	NA
Universal M13 reverse primer	CAGGAAACAGCTATGAC	55°C	NA
Universal 16S rRNA gene primer 27F	AGAGTTTGATCMTGGCTCAG	55.8°C	Lane, 1991
Universal 16S rRNA gene primer 1391R	GACGGGCGGTGTGTRCA	55.8°C	Turner *et al.*, 1999
Sedimenticola-like OTU1 1417F	AGCTAATACCGCATACGCCC	56.3°C	This chapter
Sedimenticola-like OTU1 1580R	GTGTCTCAGTCCCAGTGTGG	56.3°C	This chapter
Kistimonas-like OTU2 90F	CCTGGGAACTGCATCCCAAA	57°C	This chapter
Kistimonas-like OTU2 231R	GCACCTCAGCGTCAGTGTTA	57°C	This chapter
Spirochaeta-like OTU5 15F	GCGTTGTTCGGAATTATTGGGC	56°C	This chapter
Spirochaeta-like OTU5 226R	TCAGCGTCAATCTTTGGCCA	56°C	This chapter

Table 2.3. Features of metagenome-assembled genomes (MAGs) assembled from *P. pectinatus* gill and foot specimens.

Year of specimen collected	# reads (M)	MAG ID	Size (Mb)	No. of contigs	G+C (%)	N50 (kb)	Categorized Species	CheckM[a] completeness (%)	BUSCO[b] completeness (%)	Contamination (%)	Strain heterogeneity (%)	MAG quality[c]
2017	2.7	N1 + P5	3.0	172	63.9	26.7	*Ca.* Sedimenticola endophacoides	96.5	95.4	1.8	25.0	High
	2.7	N1 + N3 + P5	3.0	172	63.9	26.7	*Ca.* Sedimenticola endophacoides	96.5	95.4	1.8	25.0	High
2014	114	P1	2.7	368	64.5	10.2	*Ca.* Sedimenticola endophacoides	92.3	83.8	1.2	28.6	Medium
		P2	3.3	466	52.1	8.9	*Kistimonas*-like sp.	92.9	83.8	1.5	0	Medium
		P3	1.9	177	50.0	15.1	*Spirochaeta*-like sp.	86.3	70.4	0.8	0	Medium
	2.7	N3 + P5	2.9	338	64.2	11.8	*Ca.* Sedimenticola endophacoides	96.3	95.1	1.3	0	High
2013	2.1	P4	2.8	423	64.2	8.9	*Ca.* Sedimenticola endophacoides	94.4	92.5	1.2	16.7	Medium
	2.6	P5	2.7	336	64.5	11.2	*Ca.* Sedimenticola endophacoides	95.0	93.6	1.1	0	High
	1.6	P6	1.1	440	64.0	2.5	*Ca.* Sedimenticola endophacoides	38.6	26.3	1.2	0	Low
	3.0	P7	1.8	609	64.2	3.0	*Ca.* Sedimenticola endophacoides	61.9	46.2	1.4	16.7	Medium
	1.1	P8	1.4	518	64.1	2.7	*Ca.* Sedimenticola endophacoides	52.4	35.4	1.0	0	Medium
2011	1.1	P9	0.3	150	63.0	2.0	*Ca.* Sedimenticola endophacoides	20.9	7.7	0.0	0	Low
	3.0	P10	0.5	210	62.3	2.2	*Ca.* Sedimenticola endophacoides	16.1	15.9	0.0	100.0	Low
	4.9	P11	1.2	455	62.9	2.7	*Ca.* Sedimenticola endophacoides	40.1	29.0	0.6	25.0	Low
	1.6	P12	0.4	169	63.0	2.1	*Ca.* Sedimenticola endophacoides	18.9	6.9	0.0	0	Low
	4.1	P13	1.8	585	63.3	3.3	*Ca.* Sedimenticola endophacoides	58.1	50.2	1.0	12.5	Medium
	1.2	Foot	5.8	265	37.6	2.1	Unclassified	0	0	0	0	Low

[a]Parks *et al.*, 2015; [b]Simao *et al.*, 2015; [c]Bowers *et al.*, 2017

16S rRNA gene sequence analyses

Mothur v1.39.5 (Schloss *et al.*, 2009) was used for data processing for the 16S rRNA gene dataset. Operational Taxonomic Unit (OTU) clustering was performed at 99% sequence identity for higher species resolution (Nguyen *et al.*, 2016) and taxonomic classification was performed against the Silva v132 database (Quast *et al.*, 2013). The final dataset was subsampled to the library with the smallest four-digit number size (n=1,269). All reads were quality trimmed at Q=25 with Mothur v1.39.5 (Schloss *et al.*, 2009) using the trim_seq and remove_seq commands. Reads sequenced at Duke University were additionally matched to the read 1 (GCCGCGGTAA) and read 2 (GGGTNTCTAAT) primers for the V4 region of the 16S rRNA gene to exclude non-target reads and to trim off primer sequences with Mothur's pcr.seqs command. Surviving reads were processed in Mothur per the software's MiSeq SOP (Schloss *et al.*, 2009). The alignment step was modified to include the reverse complement of each sequence and report the better alignment (flip=T). Processed sequences were clustered into Operational Taxonomic Units (OTUs) and these were classified taxonomically against Silva v132 (Quast *et al.*, 2013) at 80% and 0% bootstrap confidence. Because the sequencing depths per sample were uneven, and ranged from 926 sequences to 27,323 sequences, the data was normalized by sub-sampling to 1,269 sequences (the smallest four-digit number in the dataset, eliminating two DNA samples from the 2014 collection. Using Mothur (Schloss *et al.*, 2009), the relative abundance of each OTU in each sample was computed by scaling its total abundance with the total number of sequences in the sample, and the Good's coverage estimator (Good, 1953) for each sample was calculated using the formula: C = 1 – (number

of OTUs sampled once / total number of individuals). 16S rRNA gene sequences used for phylogenetic analysis were retrieved from literature (Durand and Gros, 1996; Duperron *et al.*, 2007; Green-García, 2008; Brissac *et al.*, 2011) and from the GenBank database (Benson *et al.*, 2014) via keyword searches. All sequences were aligned with ClustalW (Thompson *et al.*, 1994) implemented in BioEdit 7.2.5 (Hall, 1999), and positions with gaps were manually trimmed off. MEGA6 (Tamura *et al.*, 2013) was used to predict the best nucleotide substitution model. The Kimura 2-parameter model (Kimura, 1980) with discrete Gamma distribution modeling of the evolutionary rate differences among sites (5 categories, with +G, parameter = 0.3382) was used to generate a maximum likelihood tree with 1,000 bootstrap replicates.

Metagenomic data analyses

Adapter removal and quality trimming (Q=30) of all Illumina-sequenced reads were performed using Cutadapt v1.11 (Martin, 2011), followed by Sickle v1.33 (Joshie and Fass, 2011). Read qualities pre- and post-trimming were assessed with FastQC v0.11.5 (Babraham Bioinformatics, 2010). Trimmed Illumina-sequenced metagenomic reads from each sequenced sample were individually assembled using IDBA-UD v1.0.9 (Peng *et al.*, 2012). Additionally, reads from the most complete gammaproteobacterial assembly were co-assembled with unprocessed Nanopore reads using the hybridSPAdes algorithm (Antipov *et al.*, 2016) of the SPAdes genome assembler (v3.11.1). For each assembly, contigs >1,500 bp long were binned with MetaBat v0.32.4 using the ensemble binning approach (Kang *et al.*, 2015), after read mapping with Bowtie2 v2.2.7 (Langmead and

Salzberg, 2012; very sensitive local and dovetail mode) and SAMtools v0.1.19 (Li *et al.*, 2009). All metagenome-assembled genomes (MAGs) were annotated with NCBI's Prokaryotic Genome Annotation Pipeline (NCBI Resource Coordinators, 2016). MAGs with >90% completeness were also annotated with Rapid Annotation using Subsystem Technology (RAST) FIGfam release 70 (Aziz *et al.*, 2008).

The quality of each metagenome-assembled genome (MAG; size, number of contigs, GC content, and Nx values) was evaluated with the Quality Assessment Tool for Genome Assemblies (QUAST, v4.5; Gurevich *et al.*, 2013). Genome lineage, completeness, contamination, and strain heterogeneity statistics were estimated by CheckM v1.0.9 (Parks *et al.*, 2015), based on a set of lineage-specific single-copy marker genes. Using reference datasets specific to each lineage identified by CheckM (Parks *et al.*, 2015), the completeness of each MAG was further evaluated with Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.1 (Simao *et al.*, 2015). Based on the completeness, contamination, presence of the 23S, 16S, and 5S rRNA genes and the number of tRNAs in each MAG, a quality measure was assigned according to the Minimum Information about a Metagenome-Assembled Genome (MIMAG) standard (Bowers *et al.*, 2017). Host (eukaryotic) genomic contamination was further assessed with MEtaGenome ANalyzer (MEGAN) community edition v6.6.4 (Huson *et al.*, 2007), which performs taxonomic assignment from web blastn (Altschul *et al.*, 1990) search results of contig sequences in each draft assembly. Ensemble binning by MetaBat (Kang *et al.*, 2015) largely removed contaminating reads from the host genome, except for the spirochete MAG where eight eukaryotic contigs were detected out of 185 contig sequences. These contigs were removed

manually from the spirochete MAG with no impact on the genome quality. Ensemble binning also failed to bin the gammaproteobacterial 16S rRNA gene sequences from the Illumina assemblies. These sequences were recovered from initial, unbinned assemblies by local blastn (Altschul *et al.*, 1990) searches with the NCBI BLAST 2.2.30+ package (NCBI Resource Coordinators, 2016) using 16S rRNA gene sequences extracted from the Nanopore co-assemblies as query. Matched contigs containing the 16S rRNA sequences were evaluated with web blastn (Altschul *et al.*, 1990) searches and added to their corresponding MAGs with no impact on the quality of these MAGs. MAG depth profiles were generated by mapping trimmed reads from each sample back to the representative, most complete *Ca.* Sedimenticola endophacoides (N1 + N3 + P5), *Kistimonas*-like sp. (P2) and *Spirochaeta*-like sp. (P3) MAGs using Bowtie2 v2.2.7 (very sensitive local and dovetail mode; Langmead and Salzberg, 2012) and SAMtools v0.1.19 (Li *et al.*, 2009). The depth of coverage of each contig in a MAG was summarized with MetaBat's v0.32.4 (Kang *et al.*, 2015) jgi_summarize_bam_contig_depths script, normalized by dividing each value the expected genome size extrapolated from CheckM (Parks *et al.*, 2015) and BUSCO (Simao *et al.*, 2015) predictions and averaged for the entire MAG.

Reciprocal average nucleotide identity (ANI) and average amino acid identity (AAI) values estimating genetic relatedness between strains and species (Konstantinidis and Tiedje, 2005) were calculated using DOE Joint Genome Institute's whole-genome based average nucleotide identity (gANI) tool (Varghese *et al.*, 2015) and CompareM v0.0.23 (Roux *et al.*, 2015), respectively. Calculated ANI and AAI values were averaged and visualized with the heatmaply package in R (https://cran.r-

project.org/web/packages/heatmaply/index.html; R Core Team, 2016). Reference genomic/MAG sequences used for phylogenomic analysis were retrieved from NCBI's GenBank (Benson *et al.*, 2014) and Genome databases (NCBI Resource Coordinators, 2016) via keyword searches. Phylogenomic analysis was performed with scripts within phylogenomics-tools (Seah, 2014), which uses AMPHORA2 (Wu and Eisen, 2008) to extract marker genes conserved in bacteria. Ten single-copy genes (*dnaG*, *frr*, *nusA*, *pgk*, *pyrG*, *rplM*, *rplS*, *rpmA*, *rpsB*, *rpsI*) present in all compared genomes/MAGs were aligned with Muscle v3.8.31 (Edgar, 2004). No marker gene was detected in the foot assembly. Each gene alignment was visually inspected to remove poorly aligned regions. The concat_align.pl script then concatenates all alignments into a partitioned alignment and predicts the best protein substitution model for each marker gene (LGF for *dnaG*, *nusA*, *pgk*; LG for *frr*, *pyrG*, *rplM*, *rplS*, *rpsB,* and *rpsI;* WAG for *rpmA*). Maximum likelihood trees with aLRT (approximate likelihood-ratio test) SH-like support values (Anisimova and Gascuel, 2006) were inferred for each partition and the combined partitions with RAxML v7.7.2 (Stamatakis, 2014). Reference protein sequences for phylogenetic classification of sulfide:quinone oxidoreductase (Sqr) and related flavocytochrome sulfide dehydrogenase (FCC) proteins were obtained from (Marcia *et al.*, 2010) and NCBI's nr (non-redundant) database (NCBI Resource Coordinators, 2016), then aligned using ClustalW (Thompson *et al.*, 1994) implemented in BioEdit 7.2.5 (Hall, 1999). MEGA v6.06 (Tamura *et al.*, 2013) was used to predict the best protein substitution model for the alignment, the LG model (Le and Gascuel, 2008), with discrete Gamma distribution modeling of the evolutionary rate differences among sites (5 categories, with +G, parameter = 5.43). Based on the model,

an unrooted maximum likelihood tree with 100 bootstrap replicates was generated. Positions with less than 95% site coverage were deleted. Classification of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) protein and hydrogenase sequences were performed via the NCBI's Basic Local Alignment Search Tool (BLAST) web tool (Altschul *et al.*, 1990) against annotated RuBisCO sequences in the nr database (NCBI Resource Coordinators, 2016) and hydrogenase sequences in the deep-sea scaly-foot snail esophageal gland symbiont (Nakagawa *et al.*, 2014). Hydrogenase class definitions in Peters *et al.* (2015) were used for classification.

Metatranscriptomic data analyses

Metatranscriptomic assembly and downstream analyses were performed with Trinity v2.5.1 (Haas *et al.*, 2013). Trimmed reads from all three metatranscriptomic libraries were co-assembled into one metatransciptome *de novo* with Trinity's default parameters (k=20). The co-assembly standardizes transcript IDs, lengths and clusters across libraries for efficient downstream quantification and cross-sample comparisons (Haas *et al.*, 2013). Trinity's Chrysalis module clusters transcripts with at least k-1 bases overlap and with sufficient reads spanning the join across both transcripts, and the Butterfly module refines the clustering and uses these transcript clusters as proxy for genes (Grabherr *et al.*, 2011; Haas *et al.*, 2013). Reads were mapped to the co-assembly using Bowtie2 v2.2.7's (Langmead and Salzberg, 2012) no-mixed, no-discordant, end-to-end options reporting up to 200 alignments per read (-k 200) and disallowing gaps within 1000 nucleotides of read extremes (-gbar 1000). Isoform and gene-level abundances were

estimated by RNA-Seq Expectation-Maximization (RSEM) that maximizes the probability of observed variables including read lengths, quality scores and sequences based on RSEM's directed graph statistical model (Li *et al.*, 2010). The probability value for each isoform/gene was divided by the effective transcript/gene length, which is the average number of possible start positions of a transcript of a given length or the abundance-weighted average effective transcript lengths of a gene's isoforms (Li *et al.*, 2010). The resulting length-normalized value for each transcript/gene was divided by the sum of length-normalized values for all transcripts/genes in each sample to calculate the transcript fraction value, which was then multiplied by 10^6 to derive the transcript per million (TPM) measure (Li *et al.*, 2010). For cross-sample comparisons, TPM values were further normalized with the trimmed means of M-values (TMM) factor that minimizes log-fold changes across samples (Robinson and Oshlack, 2010) using the edgeR Bioconductor package (Robinson *et al.*, 2010).

All assembled host and bacterial transcripts, as well as unbinned contigs from metagenomic assemblies, were annotated with Trinotate v3.1.1 (https://trinotate.github.io/), which uses the manually curated but less representative Swissprot (The UniProt Consortium, 2015) database as reference. rRNA transcripts were predicted with SortMeRNA v2 (Kopylova *et al.*, 2012) using SILVA's v119 (Quast *et al.*, 2013) collection of archaeal, bacterial, and eukaryotic 16S rRNA, 23S rRNA, 18S rRNA, and 28S rRNA gene sequences as references. Host and bacterial genes of interest were analyzed at the level of transcript clusters loosely equivalent to genes. To map transcript clusters to symbiont genes, a pan-genome for the thioautotrophic endosymbiont from *P.*

pectinatus, named *Candidatus* Sedimenticola endophacoides (explained in the Results section), was created by extracting and concatenating nucleotide sequences of RAST-annotated PEGs and RNAs from six >90% complete MAGs, followed by de-duplication with CD-HIT v4.6 (Fu *et al.*, 2012) at a global sequence identity threshold of 100%. The de-duplicated dataset was searched against the Trinity assembly using NCBI's Basic Local Alignment Search Tool (BLAST) v2.6.0+ local blastn package (Altschul *et al.*, 1990; NCBI Resource Coordinators, 2016) and only the top hit was reported (-max_target_seqs 1). Similar local blastn searches were performed on other MAGs of interest for transcript cluster to gene mapping. Functions of transcript clusters of interest were inferred by comparing Trinotate's transcript annotations with web blastp, blastn, or blastx search results (Altschul *et al.*, 1990) against the more representative NCBI's non-redundant (nr) protein sequence or nucleotide (nt) databases (NCBI Resource Coordinators, 2016) using the same 10^{-3} e-value threshold as Trinotate. For each transcript within a transcript cluster, a blastp search was performed if a likely peptide sequence was predicted by Transdecoder v5.1.0 (http://transdecoder.github.io/) based on a minimum open reading frame (ORF) length and a log-likelihood score related to the reading frame where the ORF was located. If the blastp search returned negative results or if no likely peptide sequence was predicted for a transcript, then blastn and blastx searches were performed instead. Functions of transcript clusters mapping to more than one gene were assigned based on annotations of transcript(s) within the cluster with the highest TMM-normalized TPM value(s). A transcript cluster was considered multi-mapping if more than one transcript within the

cluster shared high TMM-normalized TPM values but different predicted functions, and their corresponding genes were not adjacent to each other in the reference MAG.

qPCR and PCR

Universal primers 27F (Lane, 1991) and 1391R (Turner *et al.*, 1999) targeting the 16S rRNA gene were used to amplify total DNA extracted from one *P. pectinatus* gill sample. The amplified gene fragment was ligated into the pGEM®-T Vector System (Promega, Madison, WI, USA), followed by transformation by JM109 High-Efficiency Competent Cells (Promega) on lysogeny broth (LB)/ampicillin plates incubated at 37°C for 24 hours. White colonies were amplified with universal M13 forward and reverse primers targeting binding sites within the vector. PCR reactions were performed as previously described in this document under the following conditions: Initial denaturation at 95°C for 5 minutes, 29 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds, extension at 72°C for 30 seconds, followed by elongation at 72°C for 5 minutes and cooling at 12°C. Amplicons were run and imaged on a 1% (wt/vol) agarose gel as previously described in this document. Three transformed colonies were re-grown on another LB/ampicillin plate, re-tested with PCR and grown on liquid LB/ampicillin broth for 24 hours at 37°C. Plasmids were extracted from liquid broth using the QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA, USA), linearized with FastDigest® *Nco*I or *Nde*I (Thermo Fisher Scientific, Waltham, MA, USA) and quantified with Qubit™ dsDNA assays (Life Technologies, Austin, TX, USA). PCR-amplified gene fragments of the linearized plasmids were sequenced using the Sanger method by CUGI (Clemson, SC,

USA). Sequenced inserts compared against OTU sequences identified by 16S rRNA gene sequencing showed perfect match to the *Sedimenticola*-like OTU 1 sequence. Linearized plasmids were serially diluted and used to generate standard curves for absolute qPCR. qPCR primers for the 16S rRNA gene targeting the *Sedimenticola*-like OTU 1 (1417F and 1580R; Table 2.2) were designed using the Primer3 software (Untergasser *et al.*, 2012) integrated in Geneious v8.0 (Kearse *et al.*, 2012) based on the Sanger-sequenced insert sequences. Cloning and qPCR primers for 16S rRNA genes targeting the *Kistimonas*-like OTU2 (90F and 231R; positions based on V4 region) and *Spirochaeta*-like OTU5 (15F and 226R; positions based on V4 region; Table 2.2) were designed based on their OTU sequences using the same software. For each qPCR reaction, all samples and standards were amplified in triplicate with a 10 µl volume consisting of 0.25 µM of each primer and 1x SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad Laboratories). All qPCR amplifications were carried out in the C1000 Touch™ Thermal Cycler (Bio-Rad Laboratories) under the following conditions: Initial denaturation at 95°C for 3 minutes, 29 cycles of denaturation (33 cycles for *Spirochaeta*-like OTU5 amplification) at 95°C for 15 seconds, annealing (Table 2.2) for 30 seconds, extension at 72°C for 30 seconds, followed by elongation at 72°C for 5 minutes. Data analysis was performed with the CFX Manager software (Bio-Rad Laboratories) and all copy numbers were normalized to the amount (ng) of input DNA.

To validate nitrogen assimilation functions in the *Kistimonas*-like species, PCR primers targeting the assimilatory nitrate reductase (*nas*; 229F and 364R) and assimilatory nitrite reductase (*nit*; 454F and 580R) genes were designed (Table 2.2). To validate

vitamin B12 synthesis functions in the *Spirochaeta*-like species, PCR primers targeting the cobyrinate a,c-diamide synthase (*cbiA*; 298F and 394R) gene previously used to detect potential vitamin B12 biosynthesizers in the Ross Sea (Bertrand *et al.*, 2011) and the cobalt-precorrin-5A hydrolase (*cbiG*; 272F and 410R) genes were designed (Table 2.2). PCR primers for the *btuF* gene (692F and 794R) encoding vitamin B12-binding protein involved in transport in the *Spirochaeta*-like species were also designed (Table 2.2). PCR reactions were performed as previously described in this document under the following conditions: Initial denaturation at 95°C for 2 minutes, 33 cycles of denaturation at 95°C for 15 seconds, annealing (Table 2.2) for 30 seconds, extension at 72°C for 30 seconds, followed by elongation at 72°C for 5 minutes and cooling at 12°C. Amplicons were run and imaged on a 1% (wt/vol) agarose gel as previously described in this chapter.

Data and specimen availability

Sequence data were deposited in the National Center for Biotechnology Information (NCBI Resource Coordinators, 2016) under the BioProject ID PRJNA368737. Accession numbers are listed in Table 2.4. Dissected specimen tissues and valves are cataloged at the South Dakota School of Mines and Technology, Museum of Geology, with details provided through the iDigBio portal (https://www.idigbio.org/portal/recordsets/db3181c9-48dd-489f-96ab-a5888f5a938c).

Table 2.4. NCBI accession numbers of raw read and sequence data generated in this chapter. All data are linked to NCBI's BioProject ID PRJNA368737 (NCBI Resource Coordinators, 2016).

Database	Accession numbers	Dataset description
Sequence Read Archive (SRA)	SRR5381359-SRR5381390	16S rRNA gene sequence reads (V4 region) from the 2014 collection
	SRR6473966-SRR6473965	16S rRNA gene sequence reads (V4 region) from the 2017 collection
	SRR5381472-SRR5381483	Paired-end metagenomic reads sequenced using various Illumina platforms
	SRR6472705-SRR6472704	Metagenomic reads sequenced using the Oxford Nanopore MinION sequencing platform
	SRR6473829-SRR6473827	Metatranscriptomic reads sequenced using the Illumina HiSeq 4000 platform
GenBank	KY509297-KY509306	Nucleotide sequences of OTU1 to OTU10
	MUHZ00000000-MUIM00000000	Draft genomic assemblies from Illumina-sequenced reads
	PQCO00000000-PQCQ00000000	Draft hybrid genomic assemblies from Illumina and Oxford Nanopore reads

Results

Site characterization

Live *P. pectinatus* had clumped distributions at Wildcat Cove (Figure 2.1) in all sample years, with the highest concentrations being near the mangrove-lined coast where total organic carbon content in the sediment was highest (Doty, 2015). Overall, live abundances averaged over 40 individuals per square meter (Doty, 2015). Porewater dissolved sulfide and oxygen concentrations were measured and reported by Doty (2015) from low-flow fluid sampling of piezometers installed near to where specimens were recovered, according to previously described methods (Green-García and Engel, 2012). Dissolved sulfide concentrations at Wildcat Cove (18 – 56 μmol/L) were an order of magnitude higher than concentrations measured from intertidal zone porewater occupied by the lucinid *Lucinoma borealis* (Dando *et al.*, 1986). Dissolved oxygen concentrations ranged from 78 – 125 μmol/L at quadrats adjacent to where *P. pectinatus* were collected (Doty, 2015).

Gill microbiome diversity

To examine *P. pectinatus* microbiome diversity, we sequenced 16S rRNA genes, and used metagenomic and metatranscriptomic content from gill and foot samples collected in 2011, 2014, and 2017. All amplicon-sequenced DNA and cDNA samples were dominated by one gammaproteobacterial *Sedimenticola*-like species (OTU1), occurring at average 84 ± 11% relative abundance (Figure 2.2a). Metagenome-assembled genomes (MAGs) of this species were binned from fourteen separate assemblies and three co-assemblies (Table 2.3) and shared 100% sequence identity in the 16S rRNA gene V4 region with OTU1, as well as 99.8 ± 0.4% average nucleotide identity (ANI) and 99 ± 1% average amino acid identity (AAI; Figure 2.3a) with each other. These gammaproteobacterial MAGs were at least 20% smaller, and with at least 11% higher G+C content, than previously sequenced clade A thioautotrophic lucinid endosymbiont species *Ca.* Thiodiazotropha endoloripes (Petersen *et al.*, 2016) and *Ca.* Thiodiazotropha endolucinida (König *et al.*, 2016) and *Sedimenticola* spp. (Narasingarao and Haggblom, 2006; Flood *et al.*, 2015; Table 2.3). FISH using a newly designed SED642 probe targeting the 16S rRNA gene of this *Sedimenticola*-like species confirmed that the *P. pectinatus* gill bacteriocytes contained cells that matched the gammaproteobacterial MAGs (Figure 2.4). Results of phylogenetic analyses using 16S rRNA gene sequences (Figure 2.5) and ten single-copy marker genes (Figure 2.3b) corroborated previous reports on the distinct phylogenetic position of the thioautotrophic *P. pectinatus* endosymbiont in relation to other lucinid symbiont species (Durand *et al.*, 1996; Green-García, 2008; Brissac *et al.*, 2011). The *Sedimenticola*-like MAGs shared 71 ± 4% ANI and 64 ± 1% AAI with sequenced clade A

lucinid symbiont species, 76 ± 7% ANI and 59 ± 5% AAI with other marine thioautotrophic symbionts, and 76 ± 2% ANI and 69 ± 1% AAI to free-living *Sedimenticola* spp. (Narasingarao and Haggblom, 2006; Flood *et al.*, 2015; Figure 2.3a). Based on the 93-95% ANI and 85-90% AAI boundaries proposed in Rodriguez-R and Konstantinidis (2014), the *Sedimenticola*-like MAGs were likely a species separate from sequenced clade A lucinid symbionts, marine thioautotrophic symbionts and *Sedimenticola* spp. Because the *Sedimenticola*-like MAGs shared the highest AAI with *Sedimenticola* spp. and the observed AAI values fall within the proposed genus boundary (55-60%; Rodriguez-R and Konstantinidis, 2014), we propose the name *Candidatus* Sedimenticola endophacoides for the *P. pectinatus* endosymbiont, where 'endophacoides' refers the host association ('endo-' meaning 'within').

Besides the thioautotrophic symbiont species, we also observed lower relative abundances of a gammaproteobacterial *Kistimonas*-like OTU (average 13 ± 12%; OTU2) belonging to the order Oceanospirillales in all amplicon-sequenced DNA and cDNA samples and a *Spirochaeta*-like OTU (average 0.2 ± 0.2%; OTU5) in 25 out of 33 gill DNA and cDNA samples (Figure 2.2a). The transcriptional activity of the *Sedimenticola*-like, *Kistimonas*-like, and *Spirochaeta*-like species was confirmed by absolute qPCR quantification, where copy numbers of the OTUs in matched DNA and cDNA samples were consistent with their OTU relative abundances (Figure 2.2b).

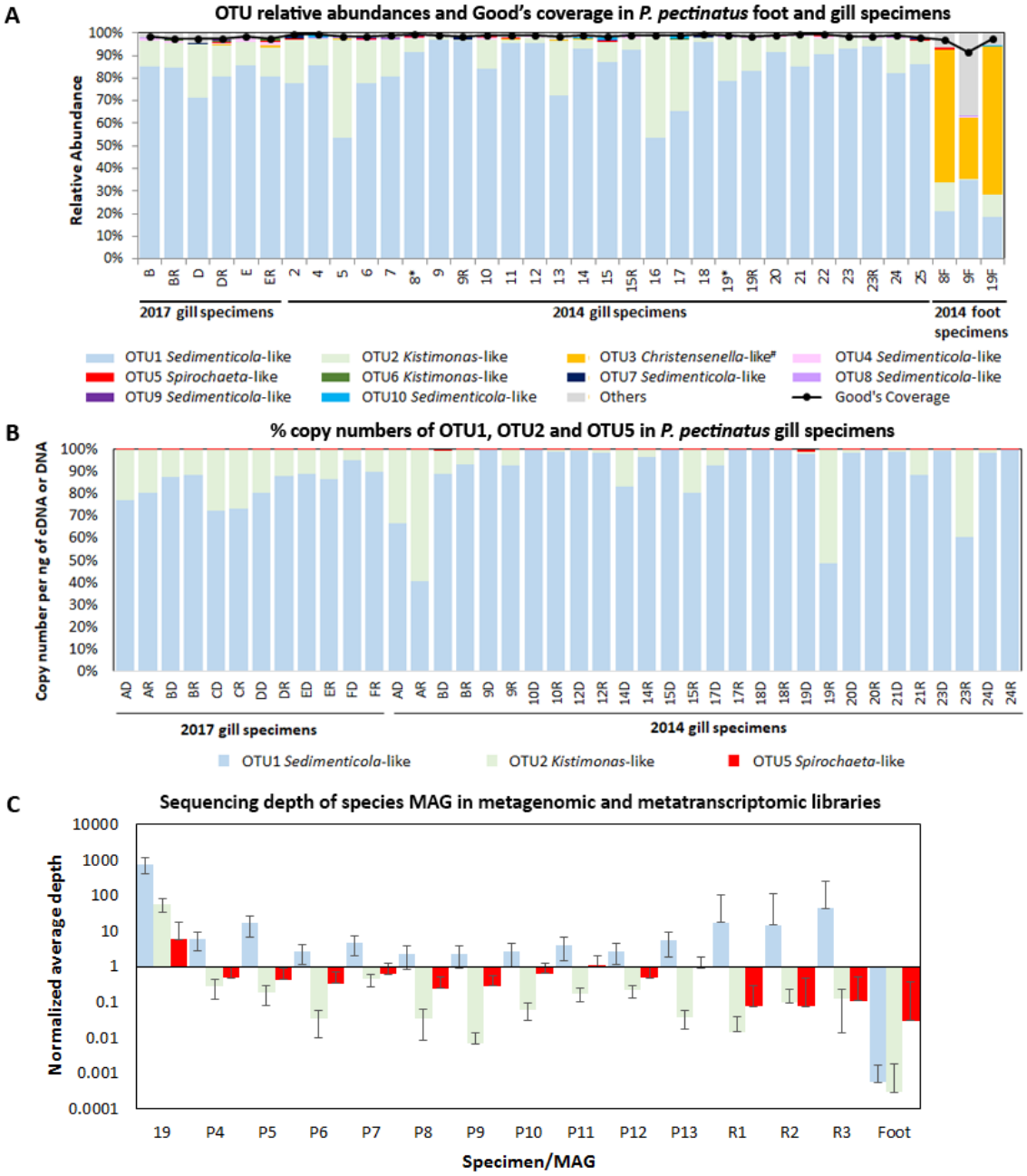

A
OTU relative abundances and Good's coverage in P. pectinatus foot and gill specimens
Relative Abundance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
B BR D DR E ER 2 4 5 6 7 8* 9 9R 10 11 12 13 14 15 15R 16 17 18 19* 19R 20 21 22 23 23R 24 25 8F 9F 19F
2017 gill specimens
2014 gill specimens
2014 foot specimens
OTU1 Sedimenticola-like
OTU2 Kistimonas-like
OTU3 Christensenella-like#
OTU4 Sedimenticola-like
OTU5 Spirochaeta-like
OTU6 Kistimonas-like
OTU7 Sedimenticola-like
OTU8 Sedimenticola-like
OTU9 Sedimenticola-like
OTU10 Sedimenticola-like
Others
Good's Coverage
B
% copy numbers of OTU1, OTU2 and OTU5 in P. pectinatus gill specimens
Copy number per ng of cDNA or DNA
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
AD AR BD BR CD CR DD DR ED ER FD FR AD AR BD BR 9D 9R 10D 10R 12D 12R 14D 14R 15D 15R 17D 17R 18D 18R 19D 19R 20D 20R 21D 21R 23D 23R 24D 24R
2017 gill specimens
2014 gill specimens
OTU1 Sedimenticola-like
OTU2 Kistimonas-like
OTU5 Spirochaeta-like
C
Sequencing depth of species MAG in metagenomic and metatranscriptomic libraries
Normalized average depth
10000
1000
100
10
1
0.1
0.01
0.001
0.0001
19 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13 R1 R2 R3 Foot
Specimen/MAG

Figure 2.2 Relative abundances of (A) bacterial 16S rRNA gene OTUs and Good's estimator of coverage (Good, 1953), (B) copy numbers per ng of DNA or cDNA (%) of *Sedimenticola*-like OTU1, *Kistimonas*-like OTU2 and *Spirochaeta*-like OTU5 determined by qPCR and (C) normalized average coverage depths with standard error bars mapped to *Ca.* Sedimenticola endophacoides, *Kistimonas*-like and *Spirochaeta*-like MAG in *P. pectinatus* foot and gill specimens/libraries. "R" denotes RNA-derived cDNA specimens in (A) and (B) and metatranscriptomic libraries in (C). Foot-associated *Christensenella*-like OTU 3 indicated with "#" in (A) was classified using 0% bootstrap confidence.

Deep metagenomic sequencing of one 2014 *P. pectinatus* gill sample also binned a *Kistimonas*-like MAG (3% of reads), a *Spirochaeta*-like MAG (0.4% of reads), a *Ca.* Sedimenticola endophacoides MAG (58% of reads; Table 2.3), and twelve other bins with 0% completeness and no taxonomic classification. These three MAGs contained 16S rRNA gene sequences with perfect matches to their corresponding OTU sequences. Unbinned contigs comprised 89% (527,385/591,741) of all assembled contigs from this sample, out of which only 11% (59,232/527,385), had predicted protein-coding regions.

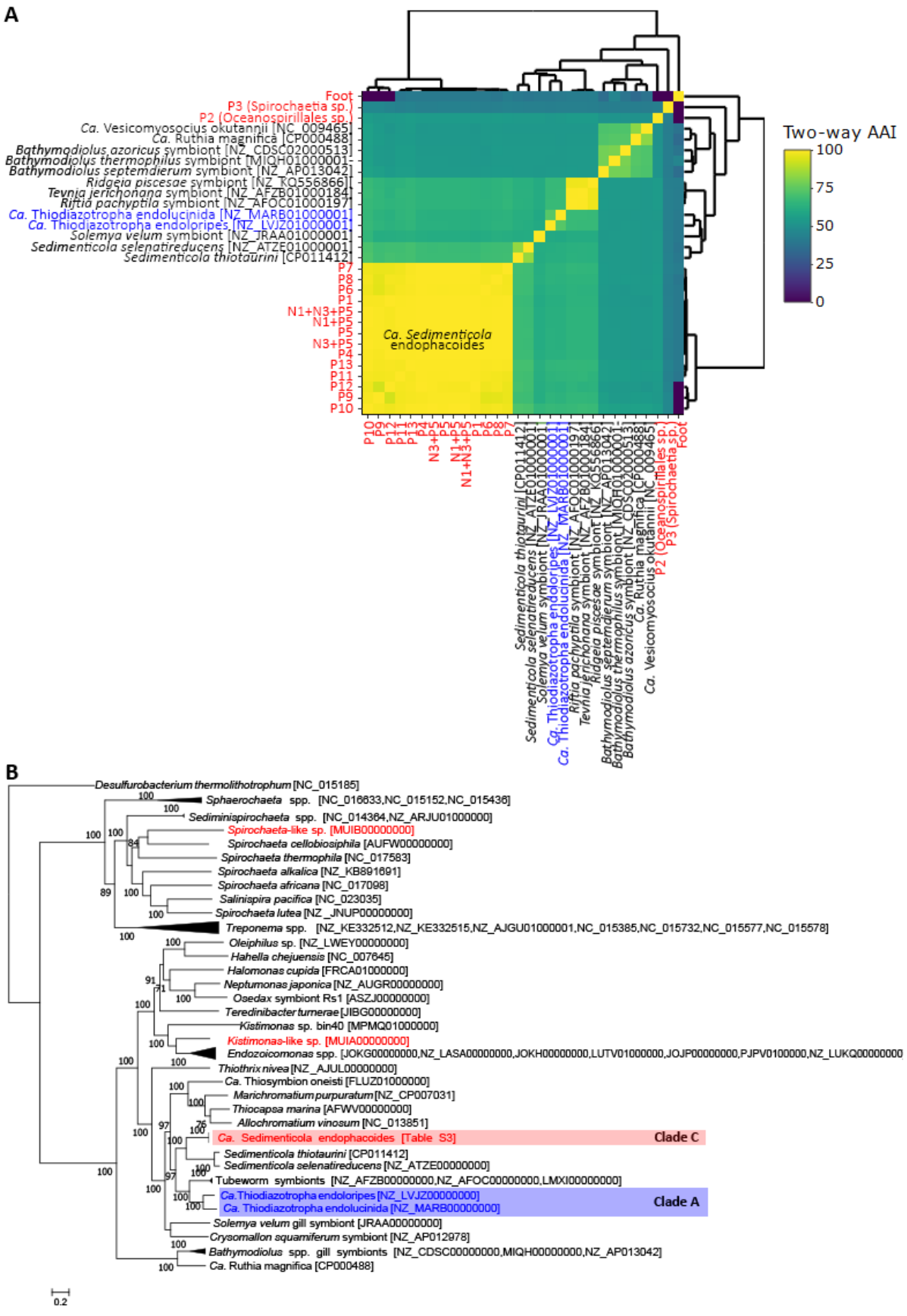

A
Foot
P3 (Spirochaetia sp.)
P2 (Oceanospirillales sp.)
Ca. Vesicomyosocius okutanii [NC_009465]
Ca. Ruthia magnifica [CP000488]
Bathymodiolus azoricus symbiont [NZ_CDSC02000513]
Bathymodiolus thermophilus symbiont [MIQH01000001]
Bathymodiolus septemdierum symbiont [NZ_AP013042]
Ridgeia piscesae symbiont [NZ_KQ556866]
Tevnia jerichonana symbiont [NZ_AFZB01000184]
Riftia pachyptila symbiont [NZ_AFOC01000197]
Ca. Thiodiazotropha endolucinida [NZ_MARB01000001]
Ca. Thiodiazotropha endoloripes [NZ_LVJZ01000001]
Solemya velum symbiont [NZ_JRAA01000001]
Sedimenticola selenatireducens [NZ_ATZE01000001]
Sedimenticola thiotaurini [CP011412]
P7
P8
P6
P1
N1+N3+P5
N1+P5
P5
N3+P5
P4
P13
P11
P12
P9
P10
Ca. Sedimenticola endophacoides
Two-way AAI
100
75
50
25
0
B
Desulfurobacterium thermolithotrophum [NC_015185]
Sphaerochaeta spp. [NC_016633,NC_015152,NC_015436]
Sediminispirochaeta spp. [NC_014364,NZ_ARJU01000000]
Spirochaeta-like sp. [MUIB00000000]
Spirochaeta cellobiosiphila [AUFW00000000]
Spirochaeta thermophila [NC_017583]
Spirochaeta alkalica [NZ_KB891691]
Spirochaeta africana [NC_017098]
Salinispira pacifica [NC_023035]
Spirochaeta lutea [NZ_JNUP00000000]
Treponema spp. [NZ_KE332512,NZ_KE332515,NZ_AJGU01000001,NC_015385,NC_015732,NC_015577,NC_015578]
Oleiphilus sp. [NZ_LWEY00000000]
Hahella chejuensis [NC_007645]
Halomonas cupida [FRCA01000000]
Neptumonas japonica [NZ_AUGR00000000]
Osedax symbiont Rs1 [ASZJ00000000]
Teredinibacter turnerae [JIBG00000000]
Kistimonas sp. bin40 [MPMQ01000000]
Kistimonas-like sp. [MUIA00000000]
Endozoicomonas spp. [JOKG00000000,NZ_LASA00000000,JOKH00000000,LUTV01000000,JOJP00000000,PJPV0100000,NZ_LUKQ00000000
Thiothrix nivea [NZ_AJUL00000000]
Ca. Thiosymbion oneisti [FLUZ01000000]
Marichromatium purpuratum [NZ_CP007031]
Thiocapsa marina [AFWV00000000]
Allochromatium vinosum [NC_013851]
Ca. Sedimenticola endophacoides [Table S3]
Clade C
Sedimenticola thiotaurini [CP011412]
Sedimenticola selenatireducens [NZ_ATZE00000000]
Tubeworm symbionts [NZ_AFZB00000000,NZ_AFOC00000000,LMXI00000000]
Ca.Thiodiazotropha endoloripes [NZ_LVJZ00000000]
Ca. Thiodiazotropha endolucinida [NZ_MARB00000000]
Clade A
Solemya velum gill symbiont [JRAA00000000]
Crysomallon squamiferum symbiont [NZ_AP012978]
Bathymodiolus spp. gill symbionts [NZ_CDSC00000000,MIQH00000000,NZ_AP013042]
Ca. Ruthia magnifica [CP000488]
0.2

Figure 2.3 (A) Heatmap of two-way, pairwise average amino acid identities (AAI) comparisons and (B) phylogenomic tree of MAGs sequenced in this study (red) in relation to published thioautotrophic symbionts of lucinid clams (blue) and other symbiotic and free-living bacteria. The outgroup used in (B) was *Desulfurobacterium thermolithotrophum* from phylum Aquificae and the scale bar indicates 0.2 substitution per site.

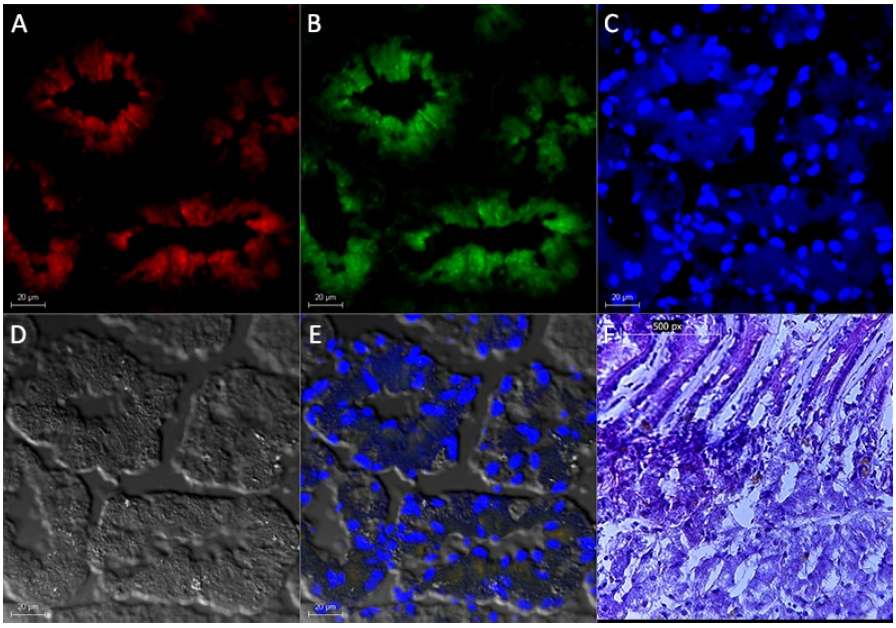


Figure 2.4. Fluorescence in situ hybridization (FISH) images of a *P. pectinatus* gill transverse section showing (A) bacteriocytes hybridized with probe SED642 specific for *Ca.* Sedimenticola endophacoides (red), (B) bacteriocytes hybridized with universal probe EUB338 (Amann *et al.*, 1990) for bacterial species (green), (C) bacteriocytes stained with DAPI (blue), (D) differential interference contrast view, (E) overlay view, (F) a light micrograph of another gill section stained with hematoxylin and eosin as a reference for tissue structural integrity and morphology.

Out of 527,385 total unbinned contigs from the deeply sequenced gill metagenome, gene/protein homologs were only predicted in 25,670 contigs (5%). ~94% of the homologs were eukaryotic, ~4% were bacterial, ~1% were viral and~ 0.2% were archaeal. ~75% of the bacterial homologs belonged to the phylum Proteobacteria, while only ~2% classified to the phylum Spirochaetes. Of all proteobacterial homologs, ~70% belonged to the class Gammaproteobacteria.

Reads from all sequenced metagenomic and metatranscriptomic libraries could be mapped to MAGs of the *Kistimonas*-like (0.4 ± 0.4% of MiSeq metagenomic reads and 0.1 ± 0.04% of metatranscriptomic reads) and *Spirochaeta*-like species (1 ± 0.3% of MiSeq metagenomic reads and 0.008 ± 0.003% of metatranscriptomic reads) at lower sequencing depths compared to the *Ca.* Sedimenticola endophacoides MAG (8 ±4% of MiSeq metagenomic reads and 1 ± 0.6% of metatranscriptomic reads; Figure 2.2c). MAGs of *Ca.* Sedimenticola endophacoides, the *Kistimonas*-like species, and the *Spirochaeta*-like species shared <70% ANI and <56% AAI with each other (Figure 2.3a). Phylogenetic analyses using 16S rRNA gene sequences clustered the *Kistimonas*-like OTU sequences with potentially pathogenic *K. scapharcae* from a dead ark clam *Anadara broughtonii* (Lee *et al.*, 2012), skin-associated *K. asteriae* from the starfish *Asterias amurensis* (Choi *et al.*, 2010), and gill-associated Oceanospirillales from the limid bivalve *Acesta excavata* (Jensen *et al.*, 2010; Figure 2.5**)**. The *Spirochaeta*-like OTU sequence was most closely related to spirochete endosymbionts in the gutless marine worm *Olavius* (Dubilier *et al.*, 1999; Blazejak *et al.*, 2005), and loosely associated with spirochete sequences retrieved

from a *L. kazani*-like lucinid (Duperron *et al.*, 2007; Figure 2.5). Genomic sequences of these closest relatives of both species are not yet available in public databases.

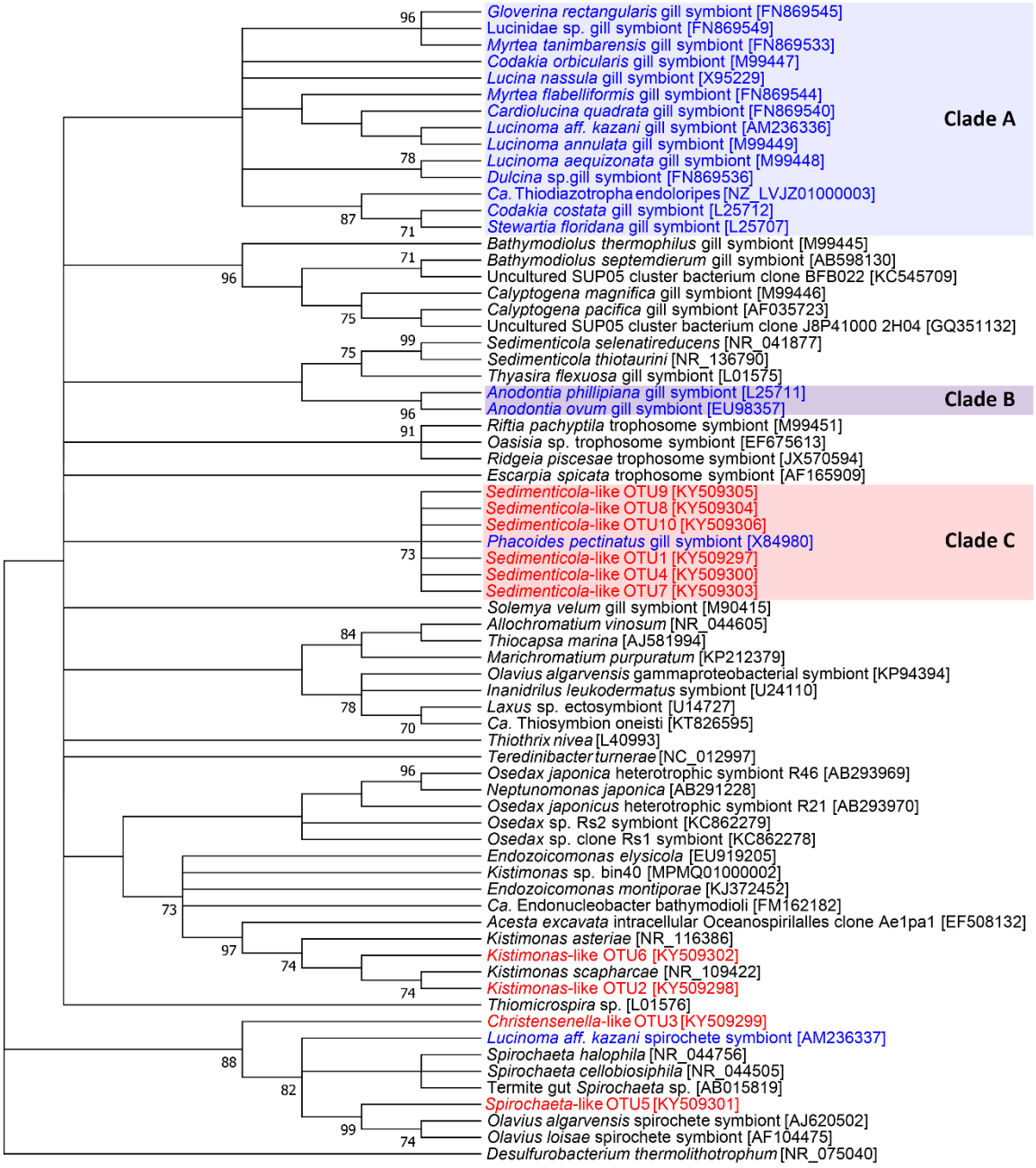


Figure 2.5. Bootstrap consensus tree of the ten most abundant 16S rRNA gene OTUs identified in this study (red text), in relation to lucinid (blue text), bivalve, tubeworm, and termite symbionts and free-living bacteria. GenBank (Benson *et al.*, 2014) accession numbers are indicated in square brackets and bootstrap values >70% are shown. The outgroup used was *Desulfurobacterium thermolithotrophum* from phylum Aquificae.

Foot microbiome diversity

Phacoides pectinatus foot DNA samples were dominated by one *Christensenella*-like OTU (OTU3) belonging to the class Clostridia at average 51 ± 21% relative abundance (Figure 2.2a). Low occurrences of the *Sedimenticola*-like OTU1 (average 25 ± 9% relative abundance), the *Kistimonas*-like OTU2 (average 8 ± 6% relative abundance) and the *Spirochaeta*-like OTU5 (average 0.5±0.3% relative abundance) were also detected in all foot specimens (Figure 2.2a). No bacterial phylogenetic marker gene or 16S rRNA gene sequence was detected in the foot metagenome. 0.02% of foot metagenomic reads mapped to the *Ca.* Sedimenticola endophacoides MAG, 0.001% mapped to the *Kistimonas*-like MAG and 2% mapped to the *Spirochaeta*-like MAG.

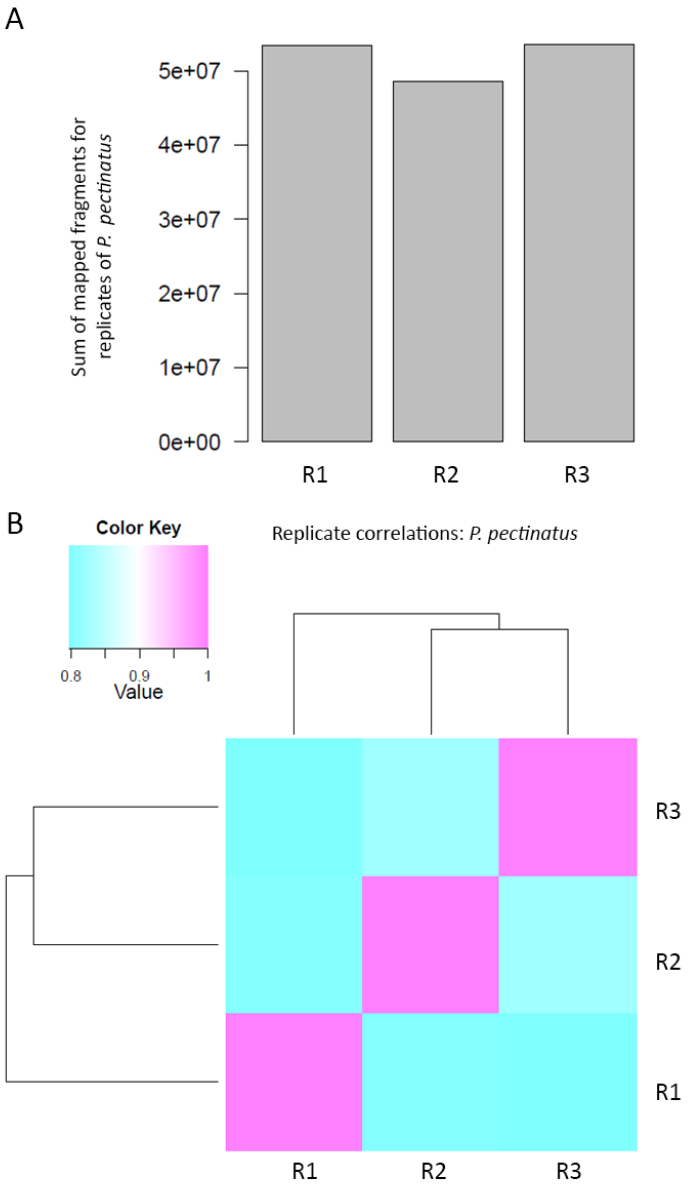


Figure 2.6. Plots showing the sum of fragments mapped to the (A) metatranscriptome assembled *de novo* by Trinity (Haas *et al.*, 2013) for each sequenced sample (R1, R2 and R3) and (B) pairwise Pearson correlations between each sequenced sample. For both plots, the count matrix was transformed to counts per million, followed by a $\log_2$ transformation.

Metagenomic and metatranscriptomic analyses

Sequenced gill cDNA libraries showed consistent read coverages of the co-assembled metatranscriptome and pairwise Pearson correlations of >0.8 across replicates (Figure 2.6). A total of 1,563,787 transcripts were assembled, out of which 85% (average length 364 ± 262 bp) were without protein-coding region and functional annotation. 57% of the 1,329,218 unclassifiable transcripts mapped to the unbinned contigs in the deeply sequenced metagenomic sample and only <0.1% of these mapped to the MAGs generated in this study. Among the 91,465 transcript clusters (loosely equivalent to genes), ~51% were eukaryote-related, ~45% were bacteria-related, ~2% were archaea-related, and ~2% were virus-related. About 2.87% of all assembled transcripts were predicted to be rRNA. 11% of the total transcripts (average length 989 ± 1,181 bp) could be mapped to gene/protein homologs. These were grouped into 91,465 transcript clusters (loosely equivalent to genes), from which a subset (3%) mapped to the bacterial MAGs of interest.

As such, it should be noted that the quality of gene/transcript annotations is heavily dependent on the completeness of the MAGs and the reference databases used. Although we made every effort to search for absent genes and pathways in the unbinned gill metagenomes, incompletely binned MAGs used to make inferences may still contain missing genes and functions. The lack of host genomic data and the high abundances of unclassifiable sequences in the gill metagenomes and metatranscriptomes imply that functional analyses can be skewed towards annotated genes/transcripts that would overlook novel genes (Raina *et al.*, 2018). Also, gene/transcript annotations based on homology may not be accurate predictors of reaction mechanisms (Raina *et al.*, 2018) and even function

(in the case of novel paralogs). Transcript quantification can also be influenced by swift changes in mRNA expression occurring between sample collection and fixation, as well as mRNA turnover that causes rapidly degrading mRNAs to exhibit inaccurately low transcripts per million (TPM) values.

Host-related functions

Host-related rRNA gene transcript clusters made up two-thirds of the thirty most abundantly expressed transcripts in the gill metatranscriptomes (Figure 2.7). Highly expressed eukaryotic and/or molluscan protein-coding genes included those encoding the respiratory cytochrome c oxidase subunits, hemoglobins 1 and 2, and actin (Figure 2.8). A carbonic anhydrase transcript cluster related to the mangrove killifish (*Kryptolebias marmoratus*) was the eleventh most abundantly expressed in the gill metatranscriptome (average 696 ± 260 TPM; Figure 2.8a), while another molluscan transcript cluster encoding for a nacrein-like protein with putative carbonic anhydrase function (Marie *et al.*, 2011) was expressed in only one out of three specimens at 0.1 TPM. The top 30 most abundant molluscan transcript clusters also included transcripts encoding hemoglobin 3 (average 104 ± 25 TPM), ribosomal proteins (average 32 ± 18 TPM), other cytoskeletal proteins (tubulin and tropomyosin; average 33 ± 19 TPM), and lysozyme 3 (average 19 ± 20 TPM; Figure 2.8b). Similarly, transcript clusters matching gene ontology (Harris *et al.*, 2004) terms relevant to hemoglobin, cytoskeletal, and ribosomal functions were among the most abundant in the phylum Mollusca (Figure 2.9). Transcript clusters involved in the defense response to bacteria (GO:0042742; Figure 2.9) were potentially relevant to symbiosis.

These included molluscan transcript clusters encoding lysozyme 1 (average 6 ± 4 TPM), lysozyme 3, an antibacterial glycoprotein aplysianin-A (Takamatsu *et al.*, 1995)/muscosal glycoprotein achacin (Ehara *et al.*, 2002; average 7 ± 2 TPM), the H_2O_2-generating flavoenzyme L-amino oxidase (Guo *et al.*, 2012; average 7 ± 1 TPM), and nitric oxide synthase (average 0.6 ± 1 TPM).

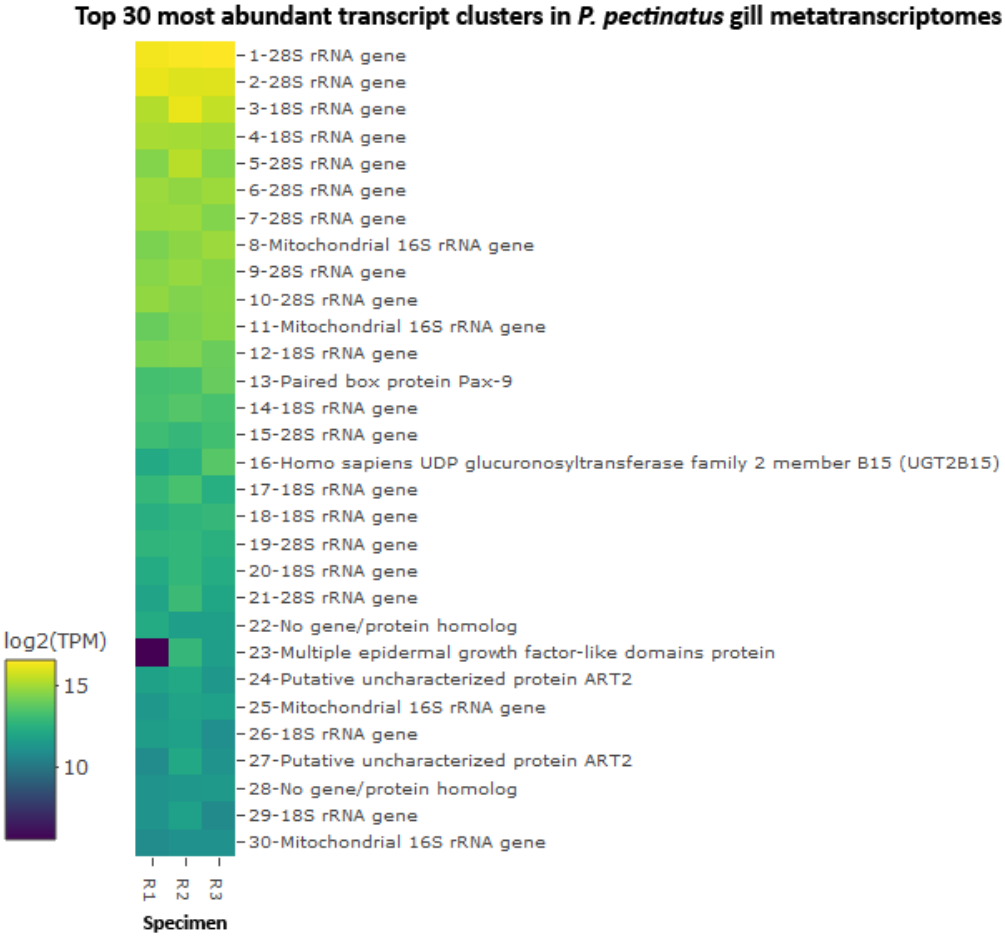


Figure 2.7. Log_2-transformed TMM-normalized TPM of gene products of the 30 most abundantly expressed transcript clusters for each sequenced metatranscriptomic sample (R1, R2 and R3) in whole *P. pectinatus* gill metatranscriptomes. Abbreviations: UDP, uridine diphosphate.

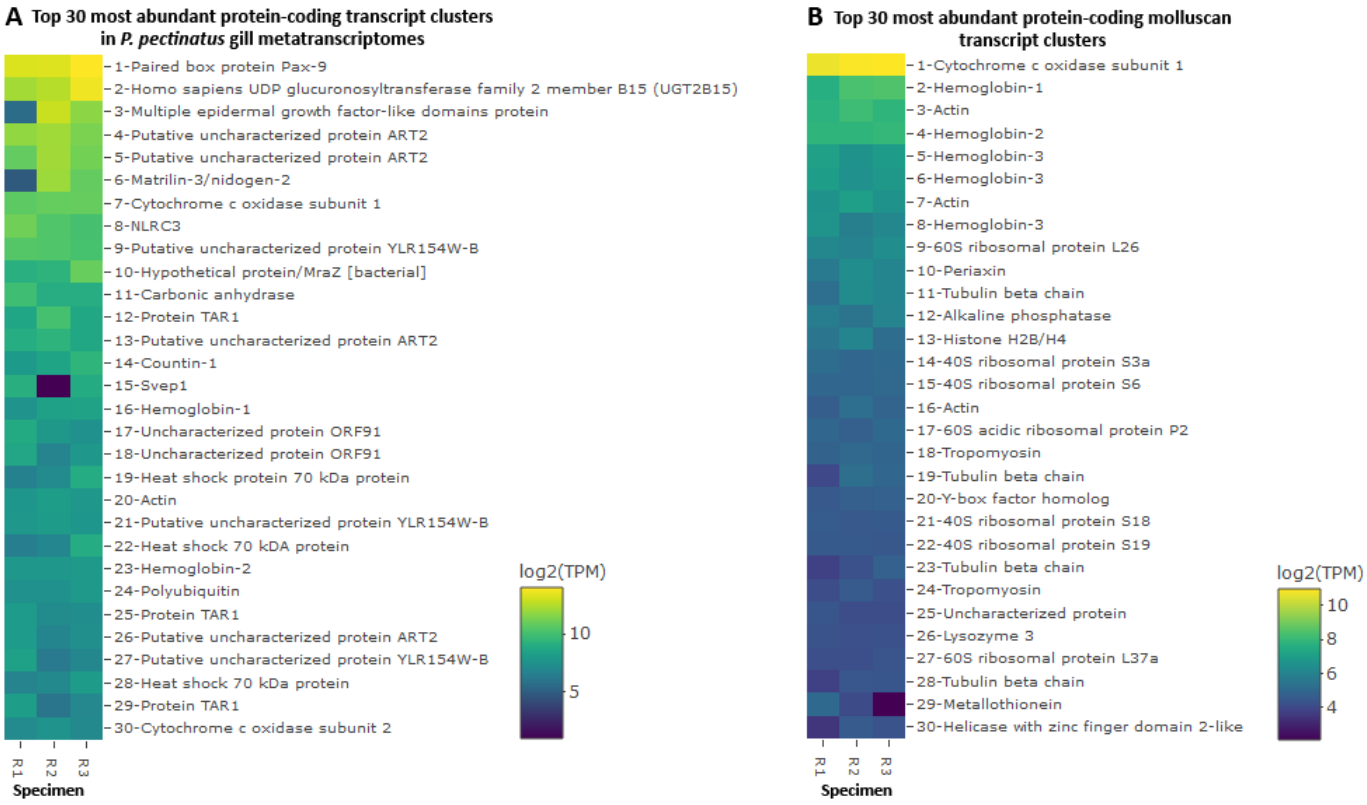


Figure 2.8 Log_2-transformed TMM-normalized TPM of gene products of the 30 most abundantly expressed protein-coding transcript clusters (A) mapped to any species and (B) mapped to the phylum Mollusca in sequenced *P. pectinatus* gill metatranscriptomes (specimens R1, R2 and R3). Abbreviations: UDP, uridine diphosphate; ORF, open reading frame.

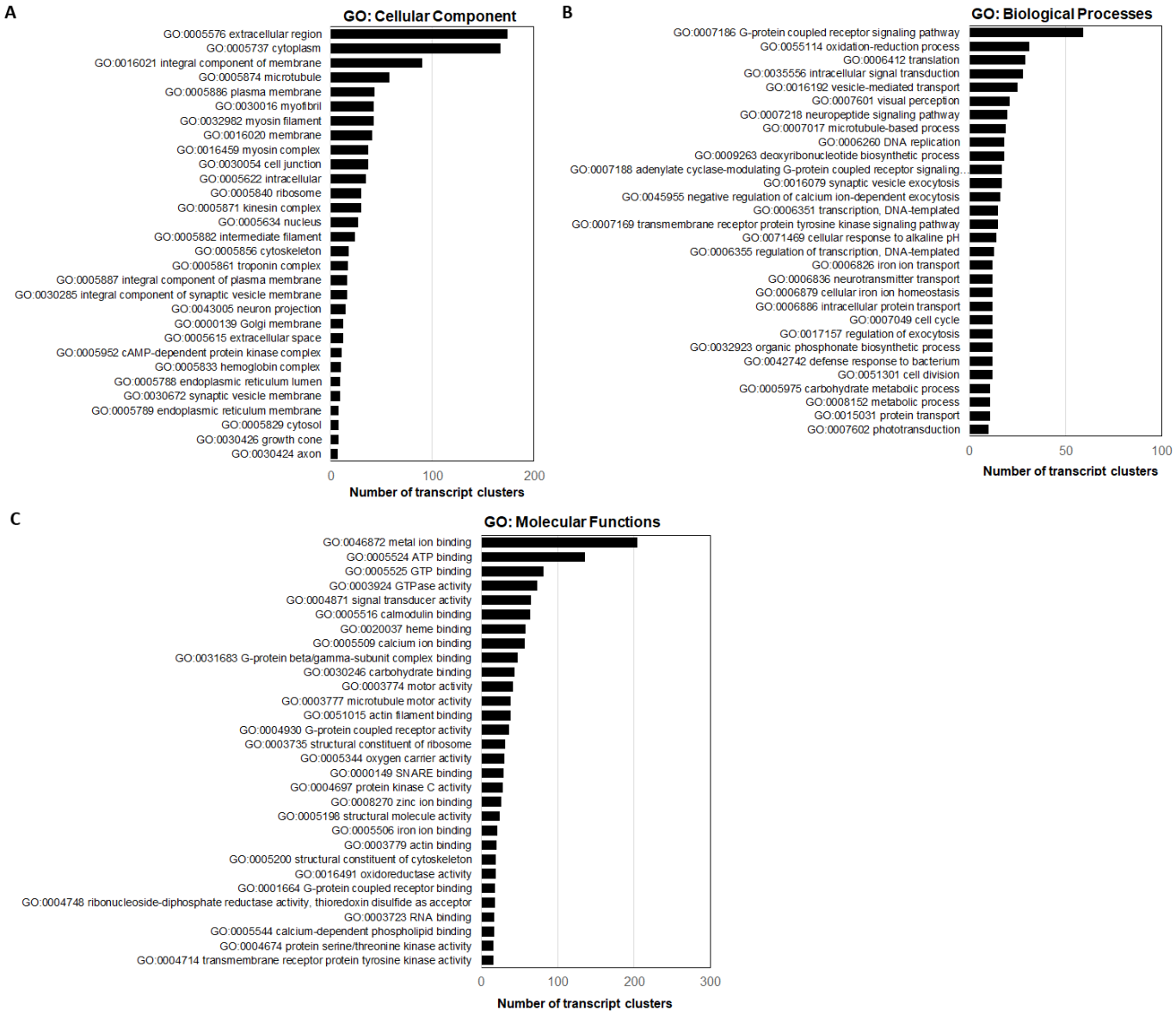


Figure 2.9. Top 30 most represented gene ontology (GO) terms (Harris *et al.*, 2004) in the (A) cellular component, (B) biological processes and (C) molecular functions categories among *P. pectinatus* transcript clusters.

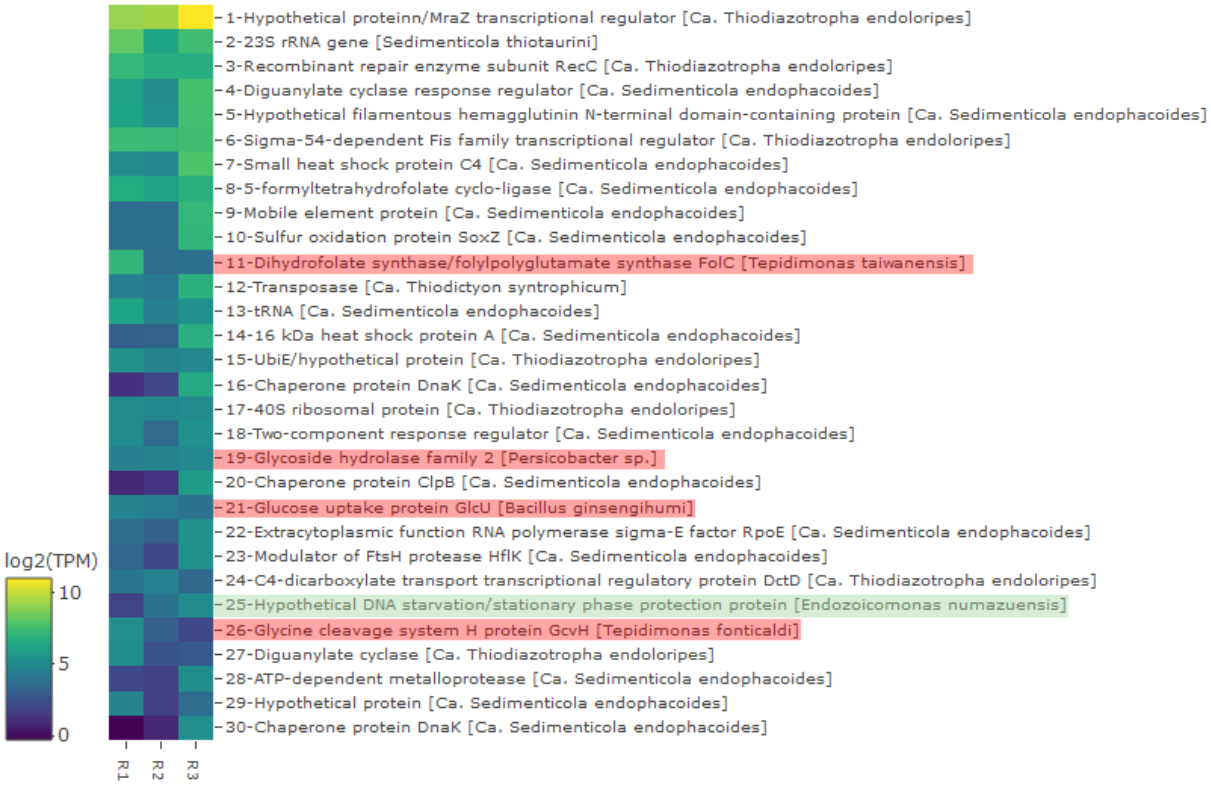
Endosymbiont functions

Sixteen of the thirty most abundant bacteria-related transcript clusters could be mapped to *Ca.* Sedimenticola endophacoides, while eight mapped to the species' relatives (Figure 2.10a). Based on Uniprot's annotations (The UniProt Consortium, 2015), many of the top 30 most abundant bacteria-related transcript clusters were involved in housekeeping functions, including DNA repair (exodeoxyribonuclease V subunit gamma RecC), transcriptional regulation (sigma-54-dependent Fis family transcriptional regulator), protein folding (filamentous temperature-sensitive ATP-dependent zinc metalloprotease FtsH and its modulator HflK), signaling (diguanylate cyclase response regulator and two-component response regulator; diguanylate cyclase), folate-dependent one-carbon metabolism or biosynthesis (5-formyltetrahydrofolate cyclo-ligase), cofactor synthesis (ubiquinone/menaquinone biosynthesis C-methyltransferase UbiE), and stress response (heat shock proteins, extracytoplasmic function RNA polymerase sigma-E factor RpoE, molecular chaperones ClpB and DnaK; Figure 2.10a).

Candidatus Sedimenticola endophacoides expressed lithoautotrophic genes involved in sulfur oxidation, hydrogen oxidation, and carbon fixation (Figures 2.10-2.12). Transcript clusters involved in thiotrophic sulfur oxidation (*sox*) and reverse dissimilatory sulfite reductase enzyme system-adenylylsulfate reductase-sulfate adenylyltransferase (*dsr-apr-sat*) pathways (Ghosh and Dam, 2009; Friedrich *et al.*, 2001) were detected in the transcriptome at TPM values between 0.07 (DsrK) to 55 (SoxZ; Figures 2.11-2.12). Variants of sulfide:quinone oxidoreductase (Sqr), hydrogenases, and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) genes utilized by chemosynthetic marine

symbionts differed across lineages (Table 2.5), and *Ca.* Sedimenticola endophacoides expressed a unique combination of type VI Sqr (average 0.09 ± 0.1 TPM), group 1 membrane-bound (average 0.2 ± 0.2 TPM) and group 2b soluble NAD-dependent (average 2 ± 2 TPM) Ni-Fe hydrogenases, and type II RuBisCO (average 0.08 ± 0.06 TPM) genes (Figures 2.11-2.12). Expressed heterotrophy-related genes included those involved in dicarboxylate transport (average 0.2 ± 0.3 TPM) and a complete TCA cycle (average 0.4 ± 0.8 TPM; Figure 2.12a). *Candidatus* Sedimenticola endophacoides is capable of respiration on oxygen and nitrogenous compounds (average 0.2 ± 0.4 TPM; Figure 2.10b). However, compared to other chemosynthetic marine symbionts that utilize a variety of terminal oxidases for aerobic respiration, we only detected genes and transcript clusters encoding subunits for the cbb3 type terminal oxidase (average 0.4 ± 0.5 TPM) in *Ca.* Sedimenticola endophacoides (Table 2.5).

A **Top 30 most abundant bacteria-related transcript clusters in *P. pectinatus* gill metatranscriptomes**



B **Major metabolic pathways predicted in *Ca.* Sedimenticola endophacoides**

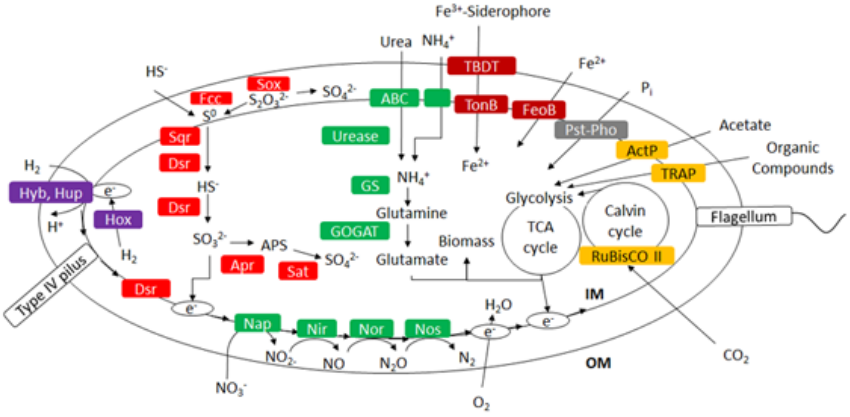


Figure 2.10. Log_2-transformed TMM-normalized TPM of gene products of (A) the 30 most abundantly expressed protein-coding transcript clusters mapped to species (square brackets) from the domain Bacteria and (B) morphological features and major metabolic pathways predicted in *Ca.* Sedimenticola endophacoides. In (B), the transcript cluster mapped to a non-thioautotrophic gammaproteobacterial species (*Endozoicomonas numazuensis*) is highlighted in green while transcript clusters mapped to non-gammaproteobacterial taxa are highlighted in pink. Abbreviations: UbiE, ubiquinone/menaquinone biosynthesis C-methyltransferase; FtsH, ATP-dependent zinc metalloprotease; Hyb, membrane bound [Ni-Fe] hydrogenase 2; Hup, uptake hydrogenase; Hox, soluble NAD-dependent hydrogenase; S^0, elemental sulfur; Fcc, Flavocytochrome c-sulfide dehydrogenase; Sqr, sulfide:quinone oxidoreductase; Sox, sulfur oxidation enzyme complex; Dsr, reverse dissimilatory sulfite reductase enzyme system; Apr, adenylylsulfate reductase; APS, adenosine-5'-phosphosulfate, Sat, sulfate adenylyltransferase; ABC, ATP-binding cassette transporters; GS, glutamine synthetase; GOGAT, glutamine oxoglutarate aminotransferase (glutamate synthase); Nap, periplasmic dissimilatory nitrate reductase; Nir, cytochrome nitrite reductase cd_1; Nor, nitric oxide reductase; Nos, nitrous oxide reductase; TBDT, TonB-dependent transporter; TonB, TonB-ExbB-ExbD complex; FeoB, ferrous iron transport protein; Pst, phosphate specific transport; Pho, phosphate regulon; PolyP, polyphosphate granule; ActP, acetate permease; TRAP, tripartite ATP-independent periplasmic transport; RuBisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; TCA cycle, tricarboxylic acid cycle; IM, inner membrane; OM, outer membrane.

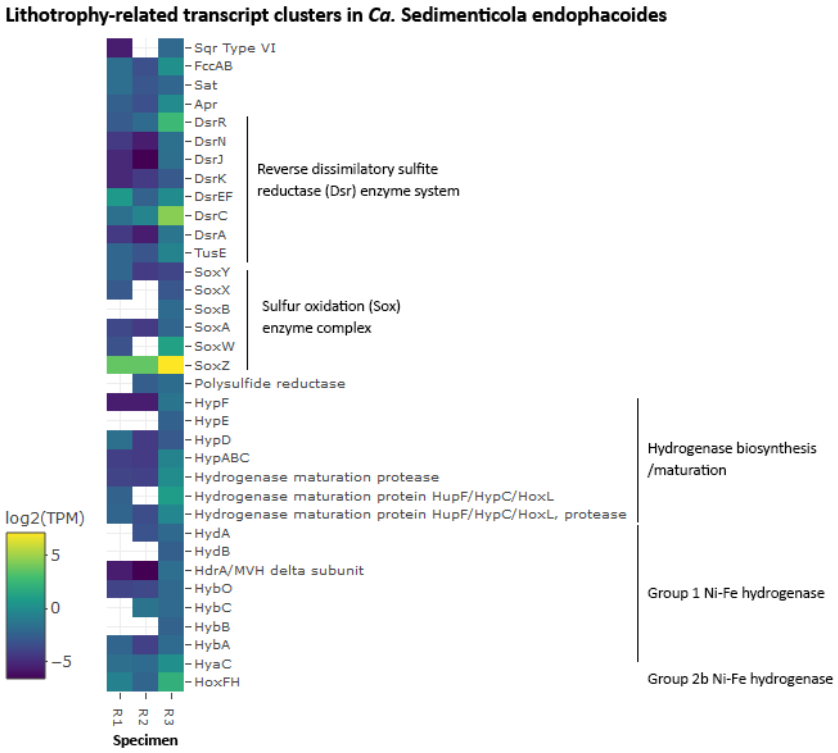


Figure 2.11. Log_2-transformed TMM-normalized TPM of gene products of lithotrophy-related transcript clusters mapped to *Ca.* Sedimenticola endophacoides. Transcript clusters with zero TPM values are represented as white cells. Abbreviations: Sqr, sulfide:quinone oxidoreductase; Fcc, flavocytochrome c – sulfide, Sat, sulfate adenylyltransferase; Apr, adenylylsulfate reductase; Tus, sulfur carrier proteins homologous to some Dsr proteins; Hyp, hydrogenase pleiotropy operon involved in the biosynthesis and maturation of [Ni-Fe] hydrogenases; Hup, regulatory uptake hydrogenase; Hox, soluble NAD-dependent hydrogenase; Hyd, periplasmic Ni-Fe hydrogenase; HdrA/MVH, heterodisulfide reductase/methylviologen reducing hydrogenase; Hyb membrane-bound Ni-Fe hydrogenase 2; HyaC, membrane-bound Ni-Fe-hydrogenase I cytochrome b subunit.

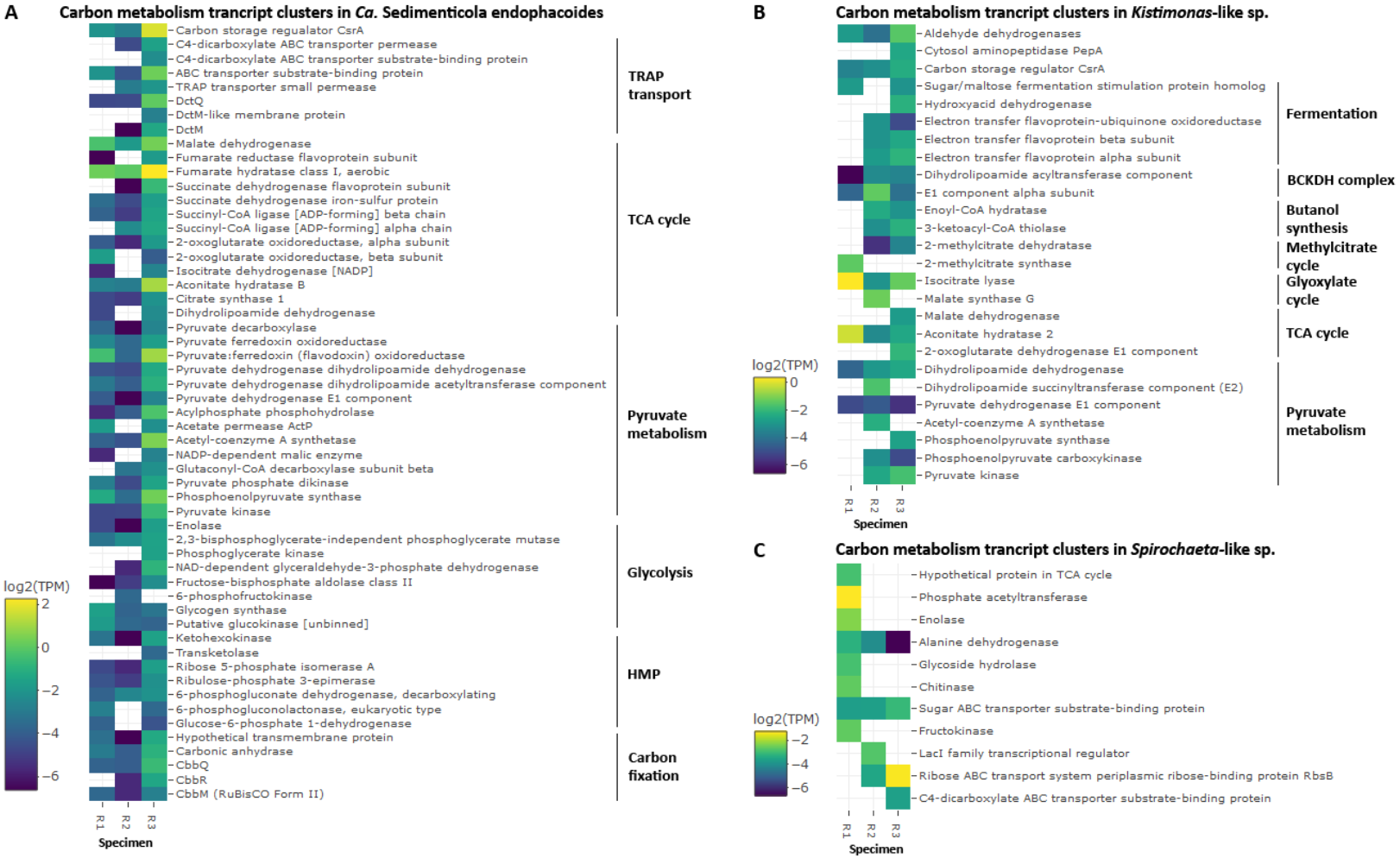

A
Carbon metabolism trancript clusters in Ca. Sedimenticola endophacoides
Carbon storage regulator CsrA
C4-dicarboxylate ABC transporter permease
C4-dicarboxylate ABC transporter substrate-binding protein
ABC transporter substrate-binding protein
TRAP transporter small permease
DctQ
DctM-like membrane protein
DctM
TRAP transport
Malate dehydrogenase
Fumarate reductase flavoprotein subunit
Fumarate hydratase class I, aerobic
Succinate dehydrogenase flavoprotein subunit
Succinate dehydrogenase iron-sulfur protein
Succinyl-CoA ligase [ADP-forming] beta chain
Succinyl-CoA ligase [ADP-forming] alpha chain
2-oxoglutarate oxidoreductase, alpha subunit
2-oxoglutarate oxidoreductase, beta subunit
Isocitrate dehydrogenase [NADP]
Aconitate hydratase B
Citrate synthase 1
TCA cycle
Dihydrolipoamide dehydrogenase
Pyruvate decarboxylase
Pyruvate ferredoxin oxidoreductase
Pyruvate:ferredoxin (flavodoxin) oxidoreductase
Pyruvate dehydrogenase dihydrolipoamide dehydrogenase
Pyruvate dehydrogenase dihydrolipoamide acetyltransferase component
Pyruvate dehydrogenase E1 component
Acylphosphate phosphohydrolase
Acetate permease ActP
Acetyl-coenzyme A synthetase
NADP-dependent malic enzyme
Glutaconyl-CoA decarboxylase subunit beta
Pyruvate phosphate dikinase
Phosphoenolpyruvate synthase
Pyruvate kinase
Pyruvate metabolism
Enolase
2,3-bisphosphoglycerate-independent phosphoglycerate mutase
Phosphoglycerate kinase
NAD-dependent glyceraldehyde-3-phosphate dehydrogenase
Fructose-bisphosphate aldolase class II
Glycolysis
6-phosphofructokinase
Glycogen synthase
Putative glucokinase [unbinned]
Ketohexokinase
Transketolase
Ribose 5-phosphate isomerase A
Ribulose-phosphate 3-epimerase
6-phosphogluconate dehydrogenase, decarboxylating
6-phosphogluconolactonase, eukaryotic type
Glucose-6-phosphate 1-dehydrogenase
HMP
Hypothetical transmembrane protein
Carbonic anhydrase
CbbQ
CbbR
CbbM (RuBisCO Form II)
Carbon fixation
log2(TPM)
2
0
−2
−4
−6
R1
R2
R3
Specimen
B
Carbon metabolism trancript clusters in Kistimonas-like sp.
Aldehyde dehydrogenases
Cytosol aminopeptidase PepA
Carbon storage regulator CsrA
Sugar/maltose fermentation stimulation protein homolog
Hydroxyacid dehydrogenase
Electron transfer flavoprotein-ubiquinone oxidoreductase
Electron transfer flavoprotein beta subunit
Electron transfer flavoprotein alpha subunit
Fermentation
Dihydrolipoamide acyltransferase component
E1 component alpha subunit
BCKDH complex
Enoyl-CoA hydratase
3-ketoacyl-CoA thiolase
Butanol synthesis
2-methylcitrate dehydratase
2-methylcitrate synthase
Methylcitrate cycle
Isocitrate lyase
Malate synthase G
Glyoxylate cycle
Malate dehydrogenase
Aconitate hydratase 2
2-oxoglutarate dehydrogenase E1 component
TCA cycle
Dihydrolipoamide dehydrogenase
Dihydrolipoamide succinyltransferase component (E2)
Pyruvate dehydrogenase E1 component
Acetyl-coenzyme A synthetase
Phosphoenolpyruvate synthase
Phosphoenolpyruvate carboxykinase
Pyruvate kinase
Pyruvate metabolism
log2(TPM)
0
−2
−4
−6
R1
R2
R3
Specimen
C
Carbon metabolism trancript clusters in Spirochaeta-like sp.
Hypothetical protein in TCA cycle
Phosphate acetyltransferase
Enolase
Alanine dehydrogenase
Glycoside hydrolase
Chitinase
Sugar ABC transporter substrate-binding protein
Fructokinase
LacI family transcriptional regulator
Ribose ABC transport system periplasmic ribose-binding protein RbsB
C4-dicarboxylate ABC transporter substrate-binding protein
log2(TPM)
−2
−4
−6
R1
R2
R3
Specimen

Figure 2.12. Log_2-transformed TMM-normalized TPM of transcript clusters encoding gene products involved in carbon metabolism mapped to (A) *Ca.* Sedimenticola endophacoides, (B) the *Kistimonas*-like species and (C) the *Spirochaeta*-like species. Transcript clusters with zero TPM values are represented as white cells. Abbreviations: TRAP, tripartite ATP-independent periplasmic transport; TCA cycle, tricarboxylic acid cycle; HMP, hexose monophosphate shunt; Dct, dicarboxylate transport proteins; Cbb, proteins encoded by the Calvin-Bassham-Benson cycle operon; BCKDH complex, branched-chain alpha-keto acid dehydrogenase complex; LacI, lactose operon repressor.

Table 2.5 Comparison of genomic features among *Ca.* Sedimenticola endophacoides, free-living *Sedimenticola* spp. (Flood *et al.*, 2015; Louie *et al.*, 2016), and bacterial symbionts, including the lucinid thioautotrophic symbiont clade A (Petersen *et al.*, 2016; Konig *et al.*, 2016), a vesicomyid clam gill symbiont (*Ca.* Ruthia Magnifica) (Roeselers *et al.*, 2010), a solemyid clam gill symbiont (Dmytrenko *et al.*, 2014), bathymodiolin mussel gill symbionts (*Bathymodiolus septemdierum* and *Bathymodiolus thermophilus*) (Ikuta *et al.*, 2016; Ponnudurai *et al.*, 2017), deep-sea tubeworm trophosomal symbionts (*Ridgeia piscesae*, *Riftia pachyptila* and *Tevnia jerichonana*) (Gardebrecht *et al.*, 2012; Perez and Juniper, 2016), a marine nematode ectosymbiont (*Ca.* Thiosymbion oneisti) (Petersen *et al.*, 2016), and a deep-sea scaly-foot snail esophageal gland symbiont (Nakagawa *et al.*, 2014). '+' denotes a feature annotated in a genome while '-' denotes a feature not yet sequenced in a genome. Abbreviations: Sqr, sulfide:quinone oxidoreductase; RuBisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase.

Function	*Ca.* Sedimenticola endophacoides	*S. thio-taurini*	*S. selenati-reducens*	*Ca.* Thio-diazotropha endoloripes	*Ca.* Thio-diazotropha endolucinida	*Ca.* Ruthia magnifica	*Solemya velum* symbiont	*Bathymodiolus* spp. symbionts	Deep-sea tubeworm symbionts	*Ca.* Thiosymbion oneisti	*Crysomallon squamiferum* symbiont
Sulfur oxidation	Sqr type VI	Sqr type I and VI			Sqr type I, III and VI	Sqr type I	Sqr type I	Sqr type I and VI			Sqr type I
Hydrogen oxidation	Group 1 and 2b (NAD-dependent) Ni-Fe hydrogenase			Group 1 and 2a Ni-Fe hydrogenase		-	Group 1 and 2b (NAD-dependent) Ni-Fe hydrogenase	Group 1 and 2a Ni-Fe hydrogenase		-	Group 1 and 2a Ni-Fe hydrogenase
Carbon fixation	RuBisCo II		RuBisCO Iaq and II	RuBisCo Iaq	RuBisCo Iaq and II	RuBisCO II	RuBisCO Iaq	RuBisCO Iaq	RuBisCO II	RuBisCO Iaq	RuBisCO Iaq and II
Nitrogen fixation	-	+	-	+	+	-	-	-	-	+	-
Assimilatory nitrate and nitrite reduction	-	-	-	+	-	-	-	+	+	+	+
Urea hydrolysis	+	+	-	+	-	-	+	-	-	+	-
Oxygen respiration	cbb3 type terminal oxidase	cbb3, aa3 and cytochrome d ubiquinol oxidases		cbb3 and aa3 type terminal oxidases							cbb3, aa3 and cytochrome d ubiquinol oxidases

Candidatus Thiodiazotropha spp. are capable of nitrogen fixation and assimilatory nitrate and nitrite reduction (Petersen *et al.*, 2016; König *et al.*, 2016), and relevant transcripts mapped to *Ca.* Thiodiazotropha endoloripes, but not *Ca.* Sedimenticola endophacoides, were identified in the gill metatranscriptomes at average 0.3 ± 0.3 TPM (Table 2.6). Though transcripts homologous to most nitrogen fixation proteins and assimilatory nitrate and nitrite reductases in *Ca.* Thiodiazotropha spp. were identified in the gill metatranscriptomes, normalized metagenomic read coverages of these transcripts were low, averaging 0 ± 0 for MiSeq sequenced reads, 0.02 ± 0.01 for Nanopore sequenced reads and 0.02 ± 0.04 for HiSeq sequenced reads (Table 2.6). In comparison, metagenomic read depths of other nitrogen metabolism transcripts averaged at 80 ± 381 for MiSeq sequenced reads, 0.009 ± 0.01 for Nanopore sequenced reads and 25,659 ± 143,903 for HiSeq sequenced reads. Key genes in these pathways were, however, not detected in the sequenced *P. pectinatus* gill metagenomes (Tables 2.6-2.7), suggesting that the transcripts were rare. MAGs of *Ca.* Sedimenticola endophacides also encoded and expressed genes for urease and the urease accessory protein UreE (0.09 ± 0.1 TPM), urea ABC transporter (0.2 ± 0.2 TPM), and ammonium transporter (average 0.06 ± 0.1 TPM; Figure 2.13).

Table 2.6. Summary of nitrogen fixation (*nif*) transcripts identified in the gill metatranscriptomes of *P. pectinatus*. The closest protein homolog and organism of each transcript were determined via bidirectional local tblastn and web tblastx searches (Altschul *et al.*, 1990) against NCBI's non-redundant protein sequences and nucleotide databases (NCBI Resource Coordinators, 2016). Trimmed means of M-values (TMM)-normalized transcripts per million (TPM) values of each sequenced metatranscriptomic sample (R1, R2 and R3) are presented with the averages and standard deviations (SD) across samples. The total average depths of coverage of each transcript for each group of metagenomic samples sequenced on MiSeq, HiSeq, and Nanopore platforms are also presented.

Transcript length	Closest protein homolog	Closest related organism	TMM-Normalized TPM					Depth		
			R1	R2	R3	Average	SD	MiSeq	HiSeq	Nanopore
751	Assimilatory nitrate reductase	*Ca.* T. endoloripes	0.3	0.1	0.1	0.2	0.1	0	0	0
2484	NAD(P)H-dependent assimilatory nitrite reductase	*Ca.* T. endoloripes	0.4	0.1	0.1	0.2	0.2	0	0.1	0.3
753	FMN-binding glutamate synthase family protein	*Ca.* T. endoloripes	0.4	0.1	0.2	0.2	0.1	0	0	0.03
4390	Homocitrate synthase, serine-O-acetyltransferase, NifWZM	*Ca.* T. endoloripes	1.1	0.2	0.9	0.7	0.4	0	0	0.02
2146	NifUS	*Ca.* T. endoloripes	0.5	0.2	0.3	0.3	0.2	0	0	0.04
4373	NifE	*Ca.* T. endoloripes	0.3	0.1	0.3	0.2	0.1	0	0.06	0.02
1125	NifNX	*Ca.* T. endoloripes	0.4	0.1	0.2	0.2	0.1	0	0	0.03
5028	NifDKT	*Ca.* T. endoloripes	0.5	0.1	0.3	0.3	0.2	0	0	0.01
1169	NifH	*Ca.* T. endoloripes	0.7	0.2	0.6	0.5	0.3	0	0	0.03

Table 2.7. Summary of local tblastn (Altschul *et al.*, 1990) search results querying translated nitrogen fixation genes in sequenced lucinid endosymbionts Ca. Thiodiazotropha endoloripes (Petersen *et al.*, 2016) and Ca. Thiodiazotropha endolucinida (König *et al.*, 2016) against unbinned P. pectinatus assemblies. Query sequences were obtained from NCBI's GenPept sequence database (NCBI Resource Coordinators, 2016). The % identity and expect value (e-value) of the best hit from each query is presented in this table.

Query	Accession Number	Organism	# Hits	% Identity	E-value	Alignment Length
Nitrogenase cofactor biosynthesis protein NifB	WP_069124667	*Ca.* T. endolucinida	1	22.69	8×10^{-5}	216
Nitrogenase iron protein	WP_069124654	*Ca.* T. endolucinida	12	22.81	7×10^{-7}	263
Nitrogen fixation negative regulator NifL	ODB98656	*Ca.* T. endoloripes	221	40.52	1×10^{-20}	116
Nitrogen fixation negative regulator NifL	ODB99078	*Ca.* T. endoloripes	211	40.52	3×10^{-20}	116
Nitrogen fixation negative regulator NifL	ODJ87650	*Ca.* T. endolucinida	208	38.98	5×10^{-19}	118
Nitrogen fixation protein VnfA	ODJ89372	*Ca.* T. endolucinida	176	42.11	1×10^{-68}	342
Nitrogen fixation protein NifM	ODJ87604	*Ca.* T. endolucinida	24	25.77	1×10^{-8}	194
Nitrogen fixation protein NifM	ODB97808	*Ca.* T. endoloripes	18	24.74	1×10^{-7}	194
Nitrogenase iron protein	WP_069005827	*Ca.* T. endoloripes	12	22.81	5×10^{-7}	263
Nitrogenase iron protein NifH	ODJ87633	*Ca.* T. endolucinida	12	22.81	7×10^{-7}	263
Nitrogen fixation protein	ODC01960	*Ca.* T. endoloripes	11	54.63	4×10^{-30}	108
Nitrogen fixation protein	ODB98544	*Ca.* T. endoloripes	11	52.78	1×10^{-28}	108
Nitrogen fixation protein FixH	ODC01716	*Ca.* T. endoloripes	11	53.99	4×10^{-61}	163
Nitrogenase cofactor biosynthesis protein NifB	WP_069014752	*Ca.* T. endoloripes	1	24.54	2×10^{-5}	216
Nitrogenase cofactor biosynthesis protein NifB	WP_069006074	*Ca.* T. endoloripes	1	24.54	2×10^{-5}	216
Nitrogenase FeMo cofactor synthesis FeS core scaffold and assembly protein NifB	ODJ87648	*Ca.* T. endolucinida	1	22.69	9×10^{-5}	216
Nitrogen fixation protein NifT	ODJ87630	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogen fixation protein NifW	ODB97810	*Ca.* T. endoloripes	0	NA	NA	NA
Nitrogen fixation protein NifX	ODB97826	*Ca.* T. endoloripes	0	NA	NA	NA
Nitrogen fixation protein NifZ	ODJ87605	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogen fixation protein NifZ	ODB98653	*Ca.* T. endoloripes	0	NA	NA	NA
Nitrogen fixation protein NifZ	ODB97809	*Ca.* T. endoloripes	0	NA	NA	NA
Nitrogenase FeMo cofactor synthesis molybdenum delivery protein NifQ	ODJ87641	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase iron-molybdenum cofactor biosynthesis protein NifE	WP_069124645	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase iron-molybdenum cofactor biosynthesis protein NifE	ODJ87623	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase iron-molybdenum cofactor biosynthesis protein NifE	WP_069024648	*Ca.* T. endoloripes	0	NA	NA	NA
Nitrogenaseiron-molybdenum cofactor biosynthesis protein NifN	WP_069124644	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase iron-molybdenum cofactor biosynthesis protein NifN	ODJ87622	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase iron-molybdenum cofactor biosynthesis protein NifN	WP_069014456	*Ca.* T. endoloripes	0	NA	NA	NA
Nitrogenase iron-molybdenum cofactor biosynthesis protein NifX	ODJ87621	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase iron-molybdenum cofactor biosynthesis protein NifY	ODJ87628	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase molybdenum-iron protein alpha chain	WP_069124653	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase molybdenum-iron protein subunit alpha NifD	ODJ87632	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase molybdenum-iron protein subunit beta	WP_069124652	*Ca.* T. endolucinida	0	NA	NA	NA
Nitrogenase molybdenum-iron protein subunit beta	WP_069005826	*Ca.* T. endoloripes	0	NA	NA	NA
Nitrogenase molybdenum-iron protein subunit beta NifK	ODJ87631	*Ca.* T. endolucinida	0	NA	NA	NA
Putative nitrogen fixation protein	ODJ87619	*Ca.* T. endolucinida	0	NA	NA	NA
Putative nitrogen fixation protein	ODJ87618	*Ca.* T. endolucinida	0	NA	NA	NA
Putative nitrogen fixation protein FixT	ODB97838	*Ca.* T. endoloripes	0	NA	NA	NA

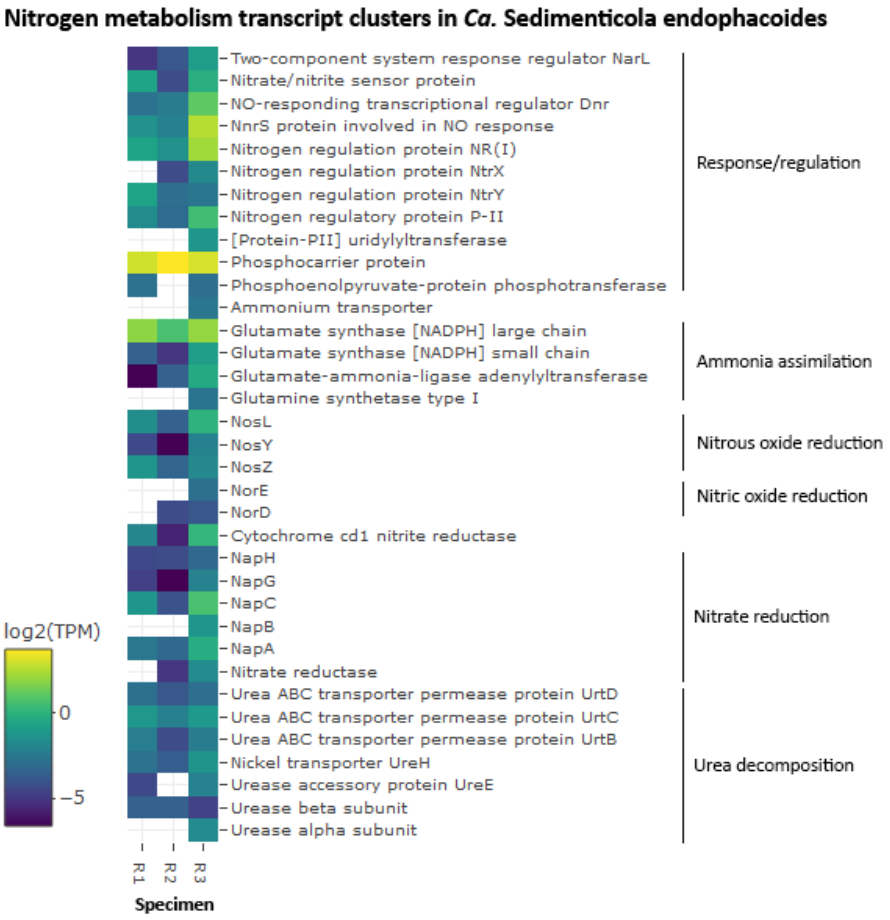


Figure 2.13. Log_2-transformed TMM-normalized TPM of transcript clusters encoding gene products involved in nitrogen metabolism mapped to *Ca.* Sedimenticola endophacoides. Transcript clusters with zero TPM values are represented as white cells. Abbreviations: Nnr, nitrite and nitric oxide reductase; Nos, nitrous oxide reductase; Nor, nitric oxide reductase; Nap, periplasmic dissimilatory nitrate reductase.

Transcripts involved in type I, II, and possibly type III and VI, secretion systems were also observed in this species (Figure 2.14). A transcript cluster encoding a hypothetical filamentous hemagglutinin N-terminal domain-containing iron-responsive protein (average 104 ± 80 TPM) secreted by the two-partner secretion system (Sun *et al.*, 2016) was also the fifth most abundant in the bacterial metatranscriptomes (Figure 2.10). Like *Ca.* Thiodiazotropha spp., *Ca.* Sedimenticola endophacoides may utilize the type I secretion system (Delepelaire, 2004) potentially for the secretion of hemolysin A (average 1 ± 1 TPM), colicin V (average 0.4 ± 0.4 TPM), and repeats in toxin (average 0.2 ± 0.03 TPM; Figure 2.14).

Although *lapBCE* genes involved in the type I secretion of the adhesin LapA were identified in MAGs, only the *lapC* transcript cluster was detected at a low average TPM of 0.2 ± 0.3 TPM. Transcript clusters encoding the general secretory pathway protein A (GspA; average 12 ± 15 TPM) and the pullulanase secretion protein E (average 0.1 ± 0.2 TPM), both of which are part of the type II secretion system, were identified. Gene homologs of type II secretion GspABCDFGHIJKLMN proteins in *Ca.* Thiodiazotropha endoloripes could only be mapped to the *P. pectinatus* gill metatranscriptomes, but not the metagenomes. These transcripts had very low metagenomic read depths, averaging 0.0004 ± 0.001 for MiSeq sequenced reads, 0.03 ± 0.01 for Nanopore sequenced reads, and 0.05 ± 0.08 for HiSeq sequenced reads. The *tadAB* and *tadD* genes, which could be part of the type IV pilus and the type II secretion system (Tomich *et al.*, 2007), were expressed in *Ca.* Sedimenticola endophacoides at average TPM of 0.3 ± 0.1, 0.06 ± 0.06, and 0.2 ± 0.2, respectively. Transcript clusters for TatABC (average 0.4 ± 0.6 TPM) and SecYEG

(average 0.2 ± 0.2 TPM) proteins within the twin-arginine translocation, and secretory protein export systems that may be linked to the type II secretion system (Nivaskumar and Francetic, 2014), were also observed in *Ca.* Sedimenticola endophacoides. Genes encoding the export apparatus protein and inner and outer membrane proteins of the type III secretion system were detected in *Ca.* Thiodiazotropha endoloripes. Protein homologs of these genes were mapped to a single 23,025 bp transcript within the *P. pectinatus* metatranscriptomes with only 0, 0.03 and 0.03 and MiSeq, Nanopore, and HiSeq metagenomic read depths. Nevertheless, transcripts encoding flagellar export proteins homologous to components of the type III secretions system (Diepold and Armitage, 2015) were identified in *Ca.* Sedimenticola endophacoides. Type VI secretion proteins annotated in *Ca.* Thiodiazotropha spp. MAGs, but not in MAGs of *Ca.* Sedimenticola endophacoides, were mapped to *P. pectinatus* metatranscriptomes with higher metagenomic read depths of 0.8 ± 3 for MiSeq sequenced reads, 0.03 ± 0.01 for Nanopore sequenced reads, and 296 ± 1,006 for HiSeq sequenced reads.

Secretion systems transcript clusters in *Ca.* Sedimenticola endophacoides

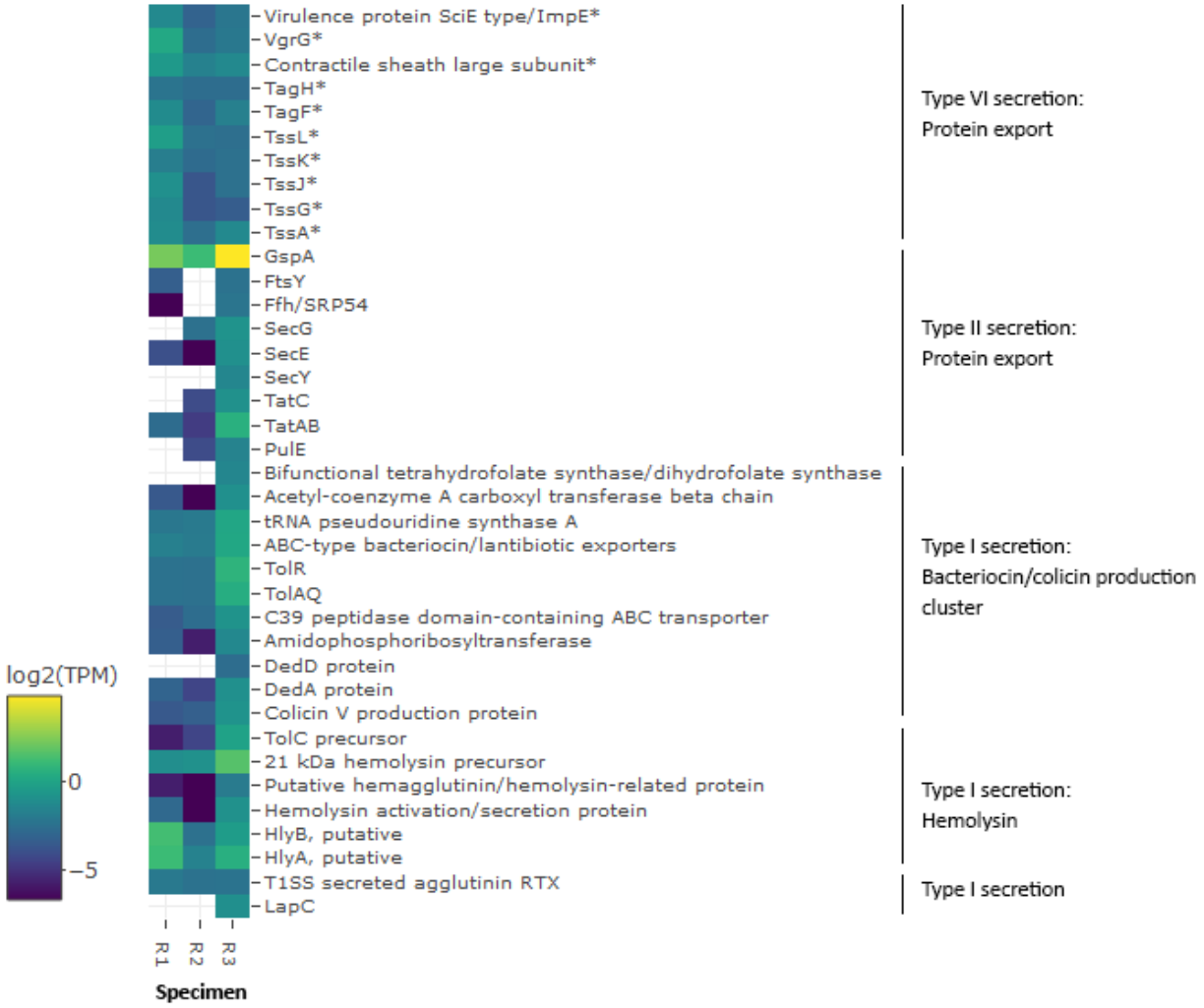


Figure 2.14. Log_2-transformed TMM-normalized TPM of transcript clusters encoding gene products involved in bacterial secretion systems mapped to *Ca.* Sedimenticola endophacoides. Transcript clusters with zero TPM values are represented as white cells. '*' indicates genes not binned in the species' MAG. Abbreviations: Imp; inner membrane protein; VgrG, valine-glycine repeat protein G; Tag, type VI secretion-associated proteins; Tss, type VI secretion system proteins; Gsp, general secretory pathway protein; FtsY, signal recognition particle receptor; Ffh/SRP 54, subunit of the signal recognition particle; Sec, secretory export proteins; Tat, twin-arginine translocation proteins; PulE, pullulanase secretion protein E; ABC, ATP-binding cassette transporters; Tol, outer membrane proteins; DedD, cell division protein; DedA; conserved ancient membrane protein; HlyB, alpha-203 hemolysin translocation ATP-binding protein; HlyA, alpha-hemolysin; T1SS, type I secretion system; RTX, repeats in toxin; LapC, large adhesion protein (membrane fusion component).

Genes to combat H_2O_2 stress, including those encoding the hydrogen peroxide-inducible genes activator (average 0.3 ± 0.3 TPM), superoxide dismutase (average 0.05 ± 0.06 TPM), and an alkyl hydroperoxide reductase subunit C-like protein (average 0.7 ± 0.8 TPM) were also expressed in *Ca.* Sedimenticola endophacoides. Motility-related genes involving the type VI pilus (average 10 ± 37 TPM), flagella (average 2 ± 4 TPM), and chemotaxis proteins (average 2 ± 2 TPM) were observed in transcriptomes of *Ca.* Sedimenticola endophacoides. Phosphate uptake was regulated via proteins encoded by the phosphate regulon (*pho*; average 0.8 ± 1 TPM) and phosphate transporter operon (*pst*; average 0.6 ± 0.9 TPM). The species could also potentially synthesize and hydrolyze inorganic polyphosphate through the activity of polyphosphate kinase and exopolyphosphatase (average 0.9 ± 1 TPM) and/or other phosphatases (average 2 ± 3 TPM). For iron scavenging, *Ca.* Sedimenticola endophacoides, like *Ca.* Thiodiazotropha spp., encoded and expressed the ferrous iron transport protein B (FeoB; average 1 ± 1 TPM), the TonB-ExbB-ExbD ferric siderophore transport system (average 0.06 ± 0.09 TPM), and the ferric uptake regulation protein FUR (average 0.2 ± 0.3 TPM). Transcript clusters for the biosynthesis of all 20 essential amino acids (Table 2.8), most B vitamins (except vitamins B3, B5 and B12; Table 2.8), coenzyme A (average 0.4 ± 0.7 TPM), tetrapyrroles (heme and siroheme; average 0.9 ± 4 TPM) and NAD and NADP cofactors (average 4 ± 9 TPM) were also identified in the species' transcriptomes.

Table 2.8. Summary of transcripts involved in amino acid and B vitamin biosynthesis in *Ca.* Sedimenticola endophacoides, the *Kistimonas*-like species and the *Spirochaeta*-like species. Amino acid names are indicated with three-letter codes. Average and standard deviation values of TMM-normalized TPM across metatranscriptomic samples are presented. "No transcript" indicates pathways present in the MAGs but not transcriptomes, while "no gene" indicates pathways not identified in the MAGs.

Category	Compound	TMM-Normalized TPM		
		***Ca.* Sedimenticola endophacoides**	***Kistimonas*-like species**	***Spirochaeta*-like species**
Branched chain amino acids	Ile, Val, Leu	0.3 ± 0.7	No gene/transcript	No transcript
Other hydrophobic amino acids	Gly	0.3 ± 0.3	No transcript	No transcript
	Ala	0.6 ± 0.9	0.06 ± 0.08	No transcript
	Pro	0.09 ± 0.1	0.2 ± 0.3	0.07 ± 0.1
	Met	0.2 ± 0.3	0.1 ± 0.2	No gene
	Trp	0.6 ± 1	0.07 ± 0.07	0.05 ± 0.08
Basic amino acids	Arg	0.2 ± 0.2	0.08 ± 0.09	No transcript
	Lys	0.2 ± 0.5	No transcript	No transcript
	His	0.3 ± 0.6	0.09 ± 2	No transcript
Acidic/polar amino acids	Glu, Glt, Asp, Asn	0.4 ± 0.8	No transcript	No transcript
	Ser	0.1 ± 0.2	0.1 ± 0.2	No transcript
	Thr	0.3 ± 0.5	0.09 ± 0.2	No transcript
	Cys	0.2 ± 0.3	0.2 ± 0.3	No transcript
Other amino acids	Phe, Tyr	0.09 ± 0.1	0.07 ± 0.01	No transcript
B vitamins	Vitamin B1	0.2 ± 0.3	No transcript	No transcript
	Vitamin B2	0.2 ± 0.3	0.07 ± 0.1	No transcript
	Vitamin B3	No gene	No gene	No gene
	Vitamin B5	No gene	No gene	No gene
	Vitamin B6	0.2 ± 0.3	0.1 ± 0.2	No transcript
	Vitamin B7	0.1 ± 0.2	0.08 ± 0.09	No transcript
	Vitamin B9	8 ± 25	0.09 ± 0.1	0.08 ± 0.1
	Vitamin B12	No gene	No gene	No transcript

Other gill microbiome functions

Highly expressed protein-coding transcript clusters homologous to protein sequences from other non-thioautotrophic bacterial taxa, including *Tepidimonas* spp., *Persicobacter* sp., and *Bacillus ginsengihumi*, were also observed in the gill metatranscriptomes (Figure 2.7a). A transcript cluster encoding a hypothetical DNA starvation/stationary phase protection protein from *Endozoicomonas numazuensis*, a relative of the *Kistimonas*-like species, was also identified (Figure 2.7a). Seven of the thirty most abundant transcript clusters mapped to the *Kistimonas*-like species encoded transposases (average 3 ± 4 TPM; Figure 2.15a). Two transcript clusters encoding poly(hydroxyalcanoate) granule associated protein (phasin) involved in the fermentative synthesis of polyhydroxyalkanoate storage granules (de Almeida *et al.*, 2007) were also highly expressed in the species (average 1 ± 1 TPM; Figure 2.15a). Heterotrophy-related genes associated with other fermentation processes were expressed by the species at lower average TPM values of 0.09 ± 0.08, along with tricarboxylic acid (TCA) cycle genes (average 0.2 ± 0.3 TPM; Figure 2.12b and Figure 2.15a). Transcript clusters linked to fatty acid catabolism and synthesis, including those involved in the glyoxylate cycle (average 0.4 ± 0.5 TPM; Munoz-Elias and McKinney, 2005), methylcitrate cycle (average 0.08 ± 0.2 TPM; Munoz-Elias *et al.*, 2006; Dolan *et al.*, 2018), and the branched-chain alpha-keto acid dehydrogenase complex (BCKDH complex; average 0.1 ± 0.1 TPM; Sun and O'Riordan, 2010; Surger *et al.*, 2018) were also observed (Figure 2.12b). A transcript cluster encoding a type VI secretion system-associated protein (average 0.5 ± 0.5 TPM) was among the most abundant in the species' transcriptomes (Figure 2.15a). The

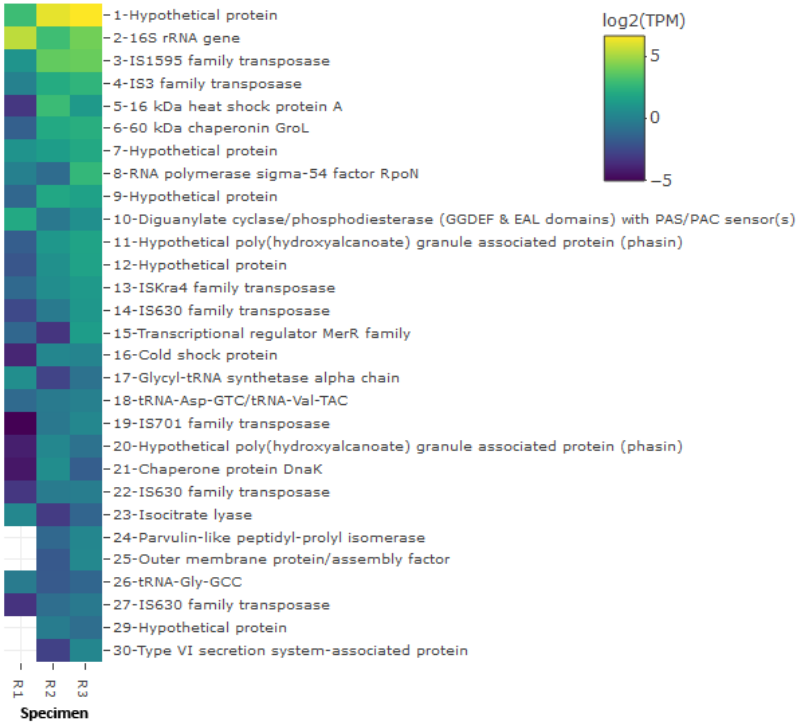
Kistimonas-like species likely respires aerobically with both cbb3 type cytochrome c oxidase (average 0.07 ± 0.07 TPM) and cytochrome bd ubiquinol oxidase (average 0.2 ± 0.3 TPM). For nitrogen assimilation (Figure 2.15c), only two genes encoding NAD(P)H-dependent assimilatory nitrite reductase (*nit*; average 0.02 ± 0.03 TPM) and type I glutamine synthetase (average 0.06 ± 0.1 TPM) were expressed in the species. Eight gill cDNA samples were amplified for the presence of assimilatory nitrate reductase (*nas*) and *nit* genes. Consistent with transcriptomic results, *nit* showed amplification in all samples tested but no amplification was detected in *nas*.

Transcriptomes of the *Kistimonas*-like species included transcript clusters involved in the transport of metal ions (sodium, potassium and cadmium; average 0.2 ± 0.3 TPM), long-chain fatty acid (average 0.2 ± 0.3 TPM), drugs (average 0.08 ± 0.1 TPM), serine (average 0.08 ± 0.1 TPM), leucine (average 0.07 ± 0.2 TPM), and other substrates (average 0.1 ± 0.1 TPM). Transcript clusters encoding the outer membrane protein OmpW (average 0.5 ± 0.4 TPM), TolC (average 0.3 ± 0.08 TPM), a type I secretion outer membrane protein (average 0.02 ± 0.04 TPM), and Omp assembly factors (average 0.1 ± 0.1 TPM) were also expressed in this species. Other transport-related functions identified in the species' MAG, but not transcriptomes, included substrates such as urea cycle products spermidine and putrescine and hemin. Bacteriocin processing genes were also detected in the species' MAG, but not transcriptomes. Like *Ca.* Sedimenticola endophacoides, the *Kistimonas*-like species expressed genes for the biosynthesis of vitamin B2, B6, B7 and B9 (Table 2.8). Vitamin B1 synthesis genes were identified in the species' MAGs, but not transcriptomes, while genes for the biosynthesis of vitamins B3, B5, and B12 were not sequenced in the

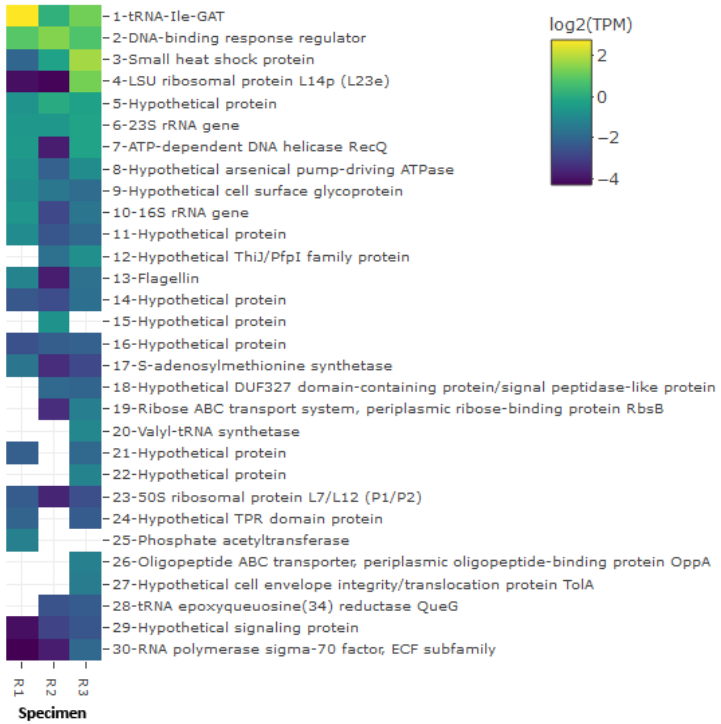
MAGs. Transcript clusters for the biosynthesis of proline (gamma-glutamyl phosphate reductase), cysteine, methionine, threonine and homoserine (homoserine dehydrogenase), serine (phosphoserine aminotransferase), histidine (phosphoribosyl-AMP cyclohydrolase), arginine, tryptophan, and alanine were detected in the bacterial species' transcriptomes (Table 2.8). On the other hand, biosynthetic genes for glycine, glutamine, glutamate, asparagine, aspartate, lysine, and leucine were identified in the MAGs, but not in the transcriptomes (Table 2.8). In the *Kistimonas*-like MAG, genes for *de novo* biosynthesis of the branched chain amino acids isoleucine and valine were not sequenced.

The most abundant transcript clusters mapped to the ~78% complete MAG of the lower abundance *Spirochaeta*-like species encoded transporters for ribose (average 0.2 ± 0.2 TPM) and oligopeptide (average 0.1 ± 0.2 TPM; Figure 2.15b). Besides ribose, the species could potentially utilize other carbon sources through transcripts encoding sugar ABC transporter substrate-binding protein (average 0.09 ± 0.02 TPM), chitinase (average 0.06 ± 0.1 TPM), glycoside hydrolase (average 0.05 ± 0.08 TPM) and C4-dicarboxylate ABC transporter substrate-binding protein (average 0.05 ± 0.08 TPM; Figure 2.12c).

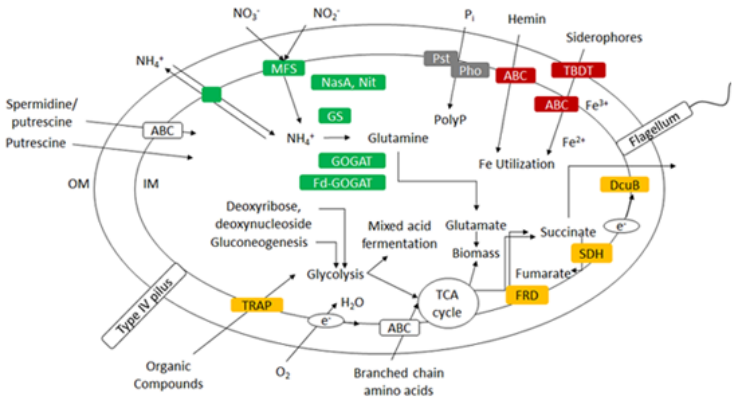
A **Top 30 most abundant transcript clusters mapped to *Kistimonas*-like sp.**



B **Top 30 most abundant transcript clusters mapped to *Spirochaeta*-like sp.**



C **Major metabolic pathways predicted in *Kistimonas*-like sp.**



D **Major metabolic pathways predicted in *Spirochaeta*-like sp.**

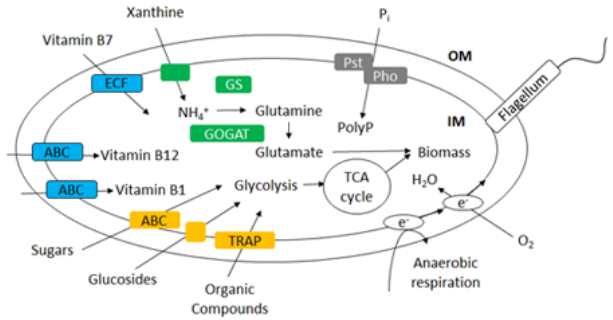


Figure 2.15. Log_2-transformed TMM-normalized TPM of gene products of the 30 most abundantly expressed protein-coding transcript clusters mapped to (A) the *Kistimonas*-like species and (B) the *Spirochaeta*-like species and major metabolic pathways predicted in (C) the *Kistimonas*-like species and (D) the *Spirochaeta*-like species. Transcript clusters with zero TPM values in (A) and (B) are represented as white cells. Abbreviations: MFS, major facilitator superfamily transporter; Nas, assimilatory nitrate reductase; Nit, assimilatory nitrite reductase; GS, glutamine synthetase; GOGAT, glutamine oxoglutarate aminotransferase (glutamate synthase); Fd-GOGAT, ferrodoxin-dependent glutamate synthase; Pst, phosphate specific transport; Pho, phosphate regulon; PolyP, polyphosphate granule; TBDT, TonB-dependent transporter; ABC, ATP-binding cassette transporters; DcuB, C4-dicarboxylate uptake family transporter; SDH, succinate dehydrogenase; FRD, fumarate reductase; TCA cycle, tricarboxylic acid cycle; TRAP, Tripartite ATP-independent periplasmic transport; ECF, energy-coupling factor transporter.

Transcript clusters for many amino acid biosynthetic pathways in the *Spirochaeta*-like species were not detected in its transcriptomes (Table 2.8), although these were predicted in its MAG. The methionine biosynthesis pathway was not sequenced in the *Spirochaeta*-like species MAG, but methionine degradation genes were expressed in its transcriptome (average 0.2 ± 0.2 TPM; Table 2.8). Genes for vitamins B1, B2, B6, B7, B9, and B12 biosynthesis were annotated in the species MAG, but only transcript clusters for vitamin B9 synthesis were observed (average 0.08 ± 0.1; Table 2.8). Genes for the transmembrane and substrate-binding component of a vitamin B1 ABC transporter, and vitamin B7 uptake proteins BioM and BioY were detected in the *Spirochaeta*-like species MAG, but not transcriptomes. The *Spirochaeta*-like MAG encoded for a nearly complete pathway (missing *cbiJ* and *cbiET*) for anaerobic vitamin B12 biosynthesis (Moore and Warren, 2012) not annotated in *Ca.* Sedimenticola endophacoides and the *Kistimonas*-like species. Although their corresponding transcripts were not detected in the species' transcriptomes, PCR targeting the cobyrinate a,c-diamide synthase (*cbiA*) and cobalt-precorrin-5A hydrolase (*cbiG*) transcripts showed amplification in 18 out of 19 gill cDNA samples and 7 out of 8 gill cDNA samples tested, respectively. We also observed PCR amplification of the *btuF* gene encoding a vitamin B12-binding protein involved in transport in eight gill cDNA samples tested. Genes for a possible nitrogen assimilation pathway with NADH-dependent glutamate synthase and purine salvage (incomplete pathway) through xanthine uptake and metabolism (Xi *et al.*, 2000) were also identified in the species' MAGs, but not transcriptomes. Respiration-related genes were not identified in both the species' MAG and transcriptomes.

Discussion

Systems-level approaches utilizing next-generation sequencing technologies successfully reveal host-microbe and microbe-microbe interactions in different invertebrate symbioses (Hansen and Moran, 2011; Rader and Nyholm, 2012; Ankrah *et al.*, 2017; Ponnudurai *et al.*, 2017), but have not been widely applied to lucinid-bacteria chemosymbioses. Currently, the lack of genomic, transcriptomic, and proteomic data for lucinids hosting gammaproteobacterial clades B and C thioautotrophic endosymbionts results in a poor understanding of the metabolism, inter- and intra- species diversity, and molecular interactions between these partners that may impact their surrounding coastal habitat and other organisms in the environment. In this study, we focused on describing the gill microbiomes of the mangrove-dwelling *P. pectinatus* that hosts the poorly characterized clade C lucinid endosymbiont species. This is the first investigation to describe the functional repertoire of 1) a lucinid symbiont species belonging to clade C, 2) a lucinid clam, and 3) other bacterial species in a lucinid gill microbiome. Our comparative genomics analyses showed thioautotrophy, respiration, and nitrogen assimilation metabolic differences among the clade C *P. pectinatus* endosymbionts, clade A lucinid symbionts, and other thioautotrophic marine symbionts, while host transcriptomes revealed candidate genes putatively involved in symbiont/microbiome selection, regulation, and nutrient transfer. Metagenomic and metatranscriptomic analyses also uncovered consistency among members of the gill microbiome, including a *Kistimonas*-like species and a *Spirochaeta*-like species that have previously been associated with a variety of marine invertebrates but not yet been comprehensively studied in lucinid clams. Additional

insights into the lucinid-bacteria chemosymbiosis is now possible, and these findings may help in species conservation, habitat management (Johnson *et al.*, 2002; Meyer *et al.*, 2008; Reynolds *et al.*, 2014), and even in fisheries productivity (Higgs *et al.*, 2016), which are areas of ongoing research.

Compared to previously sequenced lucinid clade A endosymbiont species and other thioautotrophic symbionts, *Ca.* Sedimenticola endophacoides encoded a unique combination of low affinity type VI Sqr that functions best at high sulfide concentrations (Eddie and Hanson, 2013; Shuman and Hanson, 2016), form II RuBisCO that is less efficient at discriminating between oxygen and CO_2 (Tabita *et al.*, 2008), and the high-affinity cbb3 type terminal oxidase that performs best at low oxygen concentrations (Pitcher and Watmough, 2004). These genomic differences suggest that *Ca*. Sedimenticola endophacoides experiences a more oxygen-poor extracellular and/or intracellular environment compared to *Ca.* Thiodiazotropha spp. Although pore water sulfide concentrations at Wildcat Cove were higher than previous studies (Dando *et al.*, 1986), pore water dissolved oxygen concentrations were similar to those from sub-tropical coastal mangroves (Knight *et al.*, 2013) and seagrass rhizomes (Jensen *et al.*, 2005) that have the potential to harbor lucinids. Sulfide and oxygen levels in the clam gills are likely regulated through hemoglobins, which can be partially saturated with oxygen (Wittenberg and Wittenberg, 1990). As such, sulfide-reactive hemoglobin 1, which has a higher oxygen dissociation rate than oxygen-reactive hemoglobins 2 and 3, may be confined to the symbiotic mollusc gills (Kraus and Wittenberg, 1990). In support of previous literature, we observed high expression levels of host-related hemoglobin 1, 2, and 3 genes responsible

for sulfide and oxygen transport (Kraus and Wittenberg, 1990; Frenkiel *et al.*, 1996; Rizzi *et al.*, 1996). Despite genomic evidence for the maintenance of low intracellular oxygen that would be conducive for nitrogen fixation, which can contribute to the lucinid's diet and seagrass health (van der Heide *et al.*, 2012; König *et al.*, 2016; Petersen *et al.*, 2016), *Ca.* Sedimenticola endophacoides, unlike *Ca.* Thiodiazotropha spp., is likely incapable of diazotrophy. In lieu of nitrogen fixation, we speculate that *Ca.* Sedimenticola endophacoides may utilize urea and ammonium as its nitrogen source because these transcripts were detected.

Expression levels of autotrophy-related transcripts encoding RuBisCO and Calvin cycle enzymes in relation to other transcripts were much lower for *Ca*. Sedimenticola endophacoides than previously reported in *Ca.* Thiodiazotropha endoloripes (Petersen *et al.*, 2016) and other symbiotic bivalve species that expressed RuBisCO form Iaq (Stewart *et al.*, 2011; Ponnudurai *et al.*, 2017), where these transcripts were among the most abundant in the transcriptomes. Low RuBisCO protein levels (~1%) were similarly observed in the tubeworm *Riftia pachyptila* thioautotrophic symbiont, which was discovered to produce proteins involved in an additional oxygen-sensitive reductive tricarboxylic acid cycle (Markert *et al.*, 2007; Markert *et al.*, 2011; Gardebrecht *et al.*, 2012). Although *Ca.* Sedimenticola endophacoides expressed genes encoding 2-oxoglutarate oxidoreductase that may reverse the 2-oxoglutarate to succinyl-CoA step in the TCA cycle, we did not identify any gene for citrate lyase or citryl–coenzyme A synthetase subunit that potentially converts citrate to oxaloacetate or acetate (Markert *et al.*, 2007; Gardebrecht *et al.*, 2012). Mixotrophy has previously been inferred in *Ca.*

Thiodiazotropha endoloripes (Petersen *et al.*, 2016), as well as thioautotrophic symbionts in a variety of other marine organisms (Woyke *et al.*, 2006; Dmytrenko *et al.*, 2014; Nakagawa *et al.*, 2014), and is a likely possibility for *Ca.* Sedimenticola endophacoides because of encoded and expressed genes associated with the dicarboxylate transport and TCA cycle, as well as the correlation of *P. pectinatus* live abundances to sediment organic carbon content (Doty, 2015). However, gene expression and geochemical data are insufficient support for proven mixotrophy, and more carbon assimilation experiments will be needed to determine such mechanisms in *Ca.* Sedimenticola endophacoides.

Besides *Ca.* Sedimenticola endophacoides, we also identified genes and transcripts belonging to other bacterial taxa in the *P. pectinatus* gill metagenomes and metatranscriptomes. Transcripts mapped to *Ca.* Thiodiazotropha endoloripes were noted in the gill metatransciptomes and could originate from unbinned contigs in the gill metagenomes or closely related species co-occurring in the gill microbial population. In all sequenced gill samples, we observed the consistent presence of a *Kistimonas*-like species related to the metabolically versatile Oceanospirillales species that can be symbiotic (Ding *et al.*, 2016; Neave *et al.*, 2016b; Neave *et al.*, 2016a; Schreiber *et al.*, 2016), parasitic (Zielinski *et al.*, 2009), or pathogenic (Lee *et al.*, 2012; Mendoza *et al.*, 2013). In bivalves, parasitic Oceanospirillales have been identified from nuclei in the vent mussel *Bathymodiolus* spp. (Zielinski *et al.*, 2009). Another Oceanospirillales species with unknown functions was also reported in gills from *A. excavata* (Jensen *et al.*, 2010). Consistent with previous genomic reports on Oceanospirillales species, we observed high expression of various families of transposases in the *Kistimonas*-like species, which may

facilitate rapid adaption to new hosts or environments (Katharios *et al.*, 2015; Neave *et al.*, 2017; Toshchakov *et al.*, 2017). We also identified lower relative abundances of a *Spirochaeta*-like species in most gill samples, as well as transcriptional evidence of their activity. Spirochete species have been associated with a *L. kazani*-like lucinid (Duperron *et al.*, 2007), the symbiotic gutless oligochete worm *Olavius* (Blazejak *et al.*, 2005; Ruehland *et al.*, 2008), and episymbionts of the hydrothermal vent worm *Alvinella pompejana* (Campbell and Cary, 2001).

Metatranscriptomic analyses showed that these three bacterial species may utilize distinct carbon sources. Specifically, *Ca.* Sedimenticola endophacoides may participate in mixotrophy in addition to thioautotrophy, whereas the *Kistimonas*-like species performs fermentation and fatty acid catabolism, and the *Spirochaeta*-like species breaks down chitin, sugars, and dicarboxylate compounds. To identify cellular locations of the *Kistimonas*-like and the *Spirochaeta*-like species within the host gill tissue, we designed multiple FISH probes targeting various 16S rRNA gene regions of the *Kistimonas*-like and *Spirochaeta*-like species, as these species showed positive DNA and cDNA amplification from gill specimens. However, in contrast to positive FISH signals for *Ca*. Sedimenticola endophacoides, we repeatedly failed to get unambiguous true positive signals for the *Kistimonas*-like and *Spirochaeta*-like species. This could be because of the low abundances of these species within the tissue samples, the hybridization efficiency of the designed probes, the resolution of the confocal microscopy, and/or other technical issues. Without microscopy data, we are unable to determine the location of these species and entirely rule out that they could be environmental contaminants, transient gill-filtered bacteria,

pathogens, or parasites. More sensitive techniques, such as catalyzed reporter deposition (CARD)-FISH (DeLong *et al.*, 1989) and hybridization chain reaction (HCR; Dirks and Pierce, 2004), should be performed to validate the presence of these bacteria species in the gills of *P. pectinatus*.

Our gill metatranscriptomic analyses also revealed potential host-microbiota interactions involved in the establishment and maintenance of the lucinid-bacteria relationships. In *P. pectinatus*, transfer of nutrients, including carbon and possibly B vitamins and cofactors, from symbiont to host may occur via host lysosomal digestion. The high abundances of host-associated lysozyme-encoding transcripts observed in this study may indicate the presence of active lysosomes, supporting previous reports of lysosomes in the host gills (Liberge *et al.*, 2001) and in the vent mussel *Bathymodiolus azoricus* (Ponnudurai *et al.*, 2017). We speculate that host selection may include the secretion of bactericidal lysozymes and other compounds, which can be countered by gill microbiome species. Host-related transcript clusters for the synthesis of bactericidal compounds encoding H_2O_2-generating flavoenzymes (Ehara *et al.*, 2002; Guo *et al.*, 2012), aplysianin-A (Takamatsu *et al.*, 1995), and nitric oxide (Davidson *et al.*, 2004) were weakly to moderately expressed in *P. pectinatus*. Oxidative stress-mediated symbiont selection involving nitric oxide (Davidson *et al.*, 2004) and antibacterial hypohalous acid generated from H_2O_2 and halide ions have been reported in the *Eupyrmna*-*Vibrio* symbiosis (Weis *et al.*, 1996; Small and McFall-Ngai, 1999; Schleicher and Nyholm, 2011). Gill microbiome defense to oxidative stressors likely involves weakly-expressed transcripts that detoxify hydrogen peroxide, free radicals, and nitric oxide in *Ca.* Sedimenticola endophacoides and

strongly-expressed transcripts encoding heat shock proteins in *Ca.* Sedimenticola endophacoides, and the *Kistimonas*-like and *Spirochaeta*-like species. High abundances of heat shock proteins and chaperones were also observed in the *B. azoricus* symbionts and hypothesized to be an indication of the thioautotrophic symbiont's transition into an obligate symbiont (Ponnudurai *et al.*, 2017), although other studies have described the protective functions of heat shock proteins against oxidative stress (Kalmar and Greensmith, 2009).

Presumably to decrease competition from closely-related species/strains, as speculated in the *Eupyrmna-Vibrio* symbiosis (Soto and Nishiguchi, 2014), *Ca.* Sedimenticola endophacoides encoded and expressed genes for the production and secretion of bactericidal colicin (Cascales *et al.*, 2007), which were also annotated in the *Kistimonas*-like species MAG. A strongly expressed transcript cluster encoding a hypothetical filamentous hemagglutinin N-terminal domain-containing iron-responsive protein responsible for adhesion to host tissues (Sun *et al.*, 2016) was also observed in *Ca*. Sedimenticola endophacoides, while fatty acid synthesis and catabolism-related genes encoding isocitrate lyase, BCKDH and proteins within the methylcitrate cycle in *Kistimonas*-like species have been attributed to growth and virulence in other bacterial taxa (Munoz-Elias and McKinney, 2005; Munoz-Elias *et al.*, 2006; Sun and O'Riordan, 2010; Dolan *et al.*, 2018; Surger *et al.*, 2018). Other genes associated with virulence and bacterial secretion systems were also detected in the genomes and transcriptomes of *Ca.* Sedimenticola endophacoides. However, their significance in the lucinid-bacteria chemosymbiosis is unclear. Nevertheless, the speculated roles of bactericidal, adhesion,

and virulence compounds would have to be tested using experimental approaches to better understand host selection and microbiome persistence.

Overall, this study provides insight into the metabolic functions and interactions of *P. pectinatus*, its thioautotrophic symbiont, and other gill microbiome species. Our discovery of distinct metabolic differences between the clade C endosymbiont, clade A lucinid symbionts, and other marine thioautotrophic symbionts, as well as the consistent presence and activity of other bacterial taxa in the gills, suggests that lucinid gill microbiome diversity is currently underrepresented in the literature and should warrant more investigative efforts, including additional host-microbiome meta-omics, imaging, and experimental studies. It is well established that the lucinid gill microbiome and their interactions with the host and/or the environment contribute to nutrient cycles in coastal marine sediments, however many details have been lacking. Our metagenomic and metatranscriptomic analyses of mangrove-associated lucinid host and gill microbiome functions provide a systems biology perspective of host and microbiome physiology that is relevant to host-microbe and microbe-microbe interactions.

CHAPTER THREE

EXTENSIVE THIOAUTOTROPHIC GILL ENDOSYMBIONT DIVERSITY WITHIN A SINGLE *CTENA ORBICULATA* (BIVALVIA:LUCINIDAE) POPULATION

Authors

Shen Jean Lim[1], Louie Alexander[1], Annette Summers Engel[2], Laurie C. Anderson[3], Barbara J. Campbell[1]

[1] Department of Biological Sciences, Clemson University, Clemson, SC 29634-0001

[2] Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN 37996-1410

[3] Department of Geology and Geological Engineering, South Dakota School of Mines and Technology, Rapid City, SD 57701-3901

Authors' contributions

A.S.E., B.J.C., and L.C.A. secured the funding for this study, supervised sample collection, and research efforts; S.J.L., B.J.C., A.S.E., L.C.A. collected the samples used in the study; S.J.L. and B.J.C. conceived the experiments; S.J.L. performed most of the experiments, software implementation, data analyses, and wrote this chapter; L.A. performed qPCR analyses on the thioautotrophic symbionts and assisted with bioinformatics analyses; S.J.L. maintains the NCBI sequence data and L.C.A. curates the metadata and maintains specimens of dissected tissues and valves. B.J.C reviewed and edited this chapter.

Introduction

Chemosymbiosis, where chemotrophs utilize inorganic chemical energy for the synthesis of organic compounds that benefit their hosts, is prevalent in marine bivalves, including Lucinidae clams (Dubilier *et al.*, 2008). To date, all extant lucinid bivalve species examined host chemosynthetic bacterial endosymbionts belonging to the class Gammaproteobacteria in specialized epithelial gill cells known as bacteriocytes (Taylor and Glover, 2000). Lucinid gill endosymbionts possess a diverse and varied suite of functions, including thioautotrophy (Cavanaugh *et al.*, 2006), aerobic respiration (Duplessis *et al.*, 2004b), assimilatory and dissimilatory nitrate reduction (Hentschel *et al.*, 1993; Hentschel and Felbeck, 1995; Hentschel *et al.*, 1996; König *et al.*, 2016; Petersen *et al.*, 2016), mixotrophy (Petersen *et al.*, 2016; Chapter II), hydrogenotrophy (Petersen *et al.*, 2016; Chapter II) and diazotrophy (König *et al.*, 2016; Petersen *et al.*, 2016). Consequently, the lucinid-bacteria chemosymbiosis enables lucinids to colonize habitats scarce in food, low in oxygen and high in sulfide, which are otherwise uninhabitable by other bivalves (Liljedahl, 1992). The emergence of seagrasses and mangroves during the late Cretaceous period have been associated with species radiation of shallow-marine lucinids (Stanley, 2014). Today, lucinids are commonly located near oxic-anoxic interfaces of shallow marine environments, where they burrow into anoxic sediments to obtain hydrogen sulfide and acquire oxygen from the oxic water column (Taylor and Glover, 2000; Taylor and Glover, 2010). Lucinids are frequently found in tropical and temperate seagrass beds, where free-living sulfate-reducing bacteria release hydrogen sulfide through the decomposition of dead organic matter in the sediments (Taylor and Glover, 2000;

Reynolds *et al.*, 2014). In seagrass habitats, lucinid species are thought to participate in a three-way symbiosis with their thioautotrophic gill endosymbionts and their environment (van der Heide *et al.*, 2012). Under this model, thioautotrophic lucinid endosymbionts acquire sulfide from high sulfate reduction activity in seagrass sediments to fix carbon for their host (van der Heide *et al.*, 2012). This removes toxic sulfide and possibly fixes nitrogen (Petersen *et al.*, 2016) for the seagrass beds, promoting the growth of seagrasses (van der Heide *et al.*, 2012; Reynolds *et al.*, 2014). The lucinid host, in turn, acquires oxygen for respiration from seagrass roots (van der Heide *et al.*, 2012).

Lucinid gill endosymbionts are related to a larger group of diverse marine thioautotrophic symbionts (Dubilier *et al.*, 2008). However, unlike chemosymbiotic Solemyidae and Vesicomyidae bivalves, where vertical or mixed symbiont transmission has been observed (Won *et al.*, 2003; Stewart *et al.*, 2008; Decker *et al.*, 2013), lucinid bivalves studied to date acquire their endosymbionts environmentally (Gros *et al.*, 1996; Gros *et al.*, 1998; Gros *et al.*, 1999). Based on their 16S rRNA gene sequences, lucinid endosymbionts are placed in three distinct clades, two (clades B and C) of which inhabit mangrove-dwelling lucinids and the largest (clade A) inhabit diverse seagrass-dwelling lucinids (Brissac *et al.*, 2011). While members from each clade are possibly separate species (Chapter II), clade A lucinid endosymbionts are thought to cluster as a single species with low to no variability in their 16S rRNA gene sequences (Durand and Gros, 1996; Durand *et al.*, 1996; Brissac *et al.*, 2011; Brissac *et al.*, 2016). Previously studied gill thioautotrophic endosymbionts of *Ctena orbiculata*, *Codakia orbicularis*, *Parvilucina pectinella*, *Anodontia alba*, *Divalinga quadrisulcata* and *Lucina pensylvanica* in the

Caribbean possess identical 16S rRNA gene sequences (Durand and Gros, 1996; Durand *et al.*, 1996), and their gill-symbiont fractions were capable of colonizing aposymbiotic *Codakia orbicularis* juveniles (Gros *et al.*, 2003). More recent re-analysis of these thioautotrophic symbionts using five other marker genes instead of the slow-evolving 16S rRNA gene revealed intra-specific symbiont strain diversity shaped by host geographic location (Brissac *et al.*, 2016). Strain-specific symbiont acquisition, where starved *Ctena orbiculata* individuals could only re-acquire the exact symbiont strain which they initially hosted before starvation, was also observed in the same study (Brissac *et al.*, 2016).

Besides marker-gene based diversity studies (Brissac *et al.*, 2016), -omics approaches have been used to characterize functions of clade A lucinid endosymbionts, including the *Codakia orbicularis* symbiont *Ca.* Thiodiazotropha endolucinida (König *et al.*, 2016) and the *Loripes orbiculata* symbiont *Ca.* Thiodiazotropha endoloripes (Petersen *et al.*, 2016). However, -omics data generated from these studies were not applied to investigate symbiont taxonomic, genetic and functional diversity within their study populations. Similar -omics approaches were also used to study gill microbiome diversity in a mangrove-dwelling *Phacoides pectinatus* population hosting a clade C endosymbiont (Chapter II). Approaches like the latter could potentially reveal gill microbiome and symbiont diversity in clade A symbionts at a finer resolution, which would allow useful inter-host and inter-population comparisons and provide new insights on host-symbiont specificity and possibly spatial or environmental drivers of symbiont diversity.

In this study, we focused on characterizing the taxonomic, genetic and functional composition of symbiont communities within *Ctena orbiculata* (Montagu, 1808)

individuals dominating a mixed lucinid population at Sammy Creek Landing, Sugarloaf Key, USA. Besides strain-level symbiont diversity, we also sought to investigate possible influences of spatial and environmental factors on symbiont diversity and host-symbiont functions by comparing gill samples from seagrass-covered quadrats and algae-covered quadrats. To this end, we sequenced the gill microbiomes and metagenomes of *C. orbiculata* to generate bacterial taxonomic profiles and metagenome-assembled genomes (MAGs). We then performed metatranscriptomic analyses to infer host-symbiont gene expression and to identify differentially expressed genes across taxa and quadrats.

Materials and methods

Sample collection

Ctena orbiculata clams were collected from Sammy Creek Landing, Sugarloaf Key, Florida in June 2016, within quadrats set up along two 50 m transects (T20 and T21) at 5 m intervals perpendicular to the shoreline (Figure 3.1). A third transect (T22) parallel to the first two transects was also set up to sample *C. orbiculata* bivalves in a quadrat 3.5 m away from shore containing 100% vegetation coverage of algae. 1 m^2 quadrats were used for the transect T20 while 0.25 m^2 quadrats were used for transects T21 and T22. Specimens were sieved from sediments excavated to a layer rich in microfloral debris that demarked base of active bioturbation. Collected specimens were preserved in RNAlater within 30 minutes of collection. Tissue dissection, nucleic acids extraction, cDNA synthesis, and fluorometric quantification steps were performed as described in Chapter II.

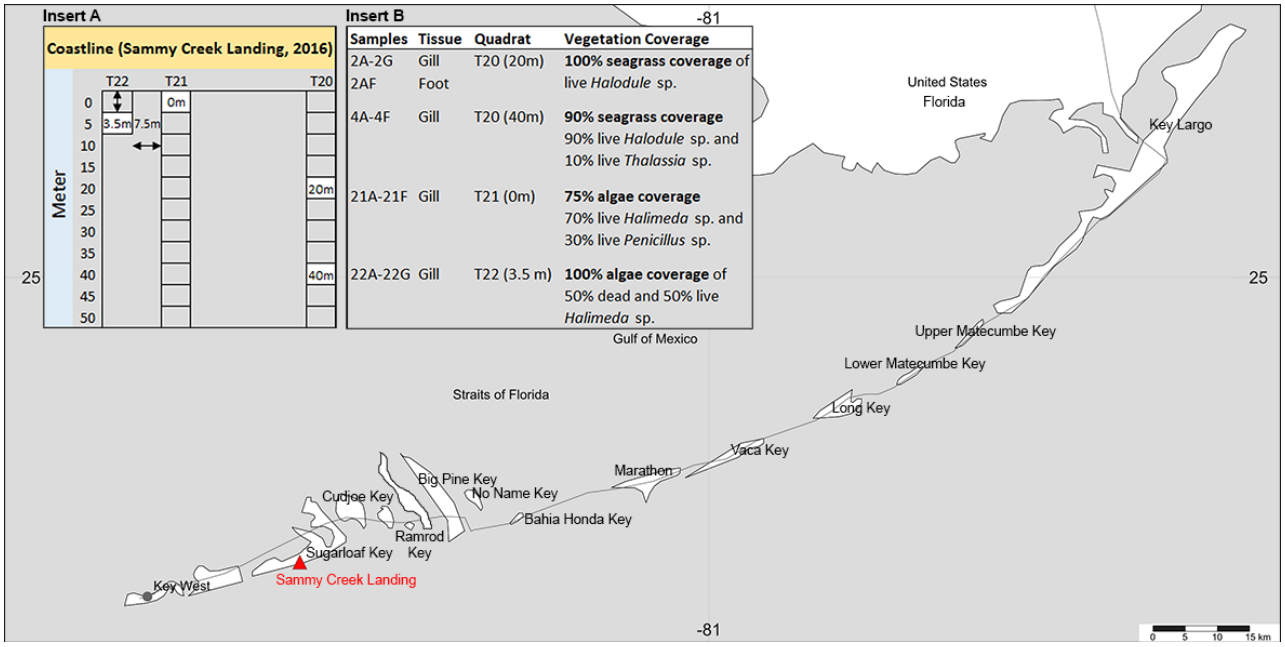


Figure 3.1. Map showing location of the sampling site (Sammy Creek Landing) in Florida, USA. *Ctena orbiculata* specimens were collected from four quadrats (white cells) with varying distances to the shoreline within three transects (T20, T21 and T22; Insert A). Insert B shows specimen and vegetation coverage information of each quadrat within the collection site.

16S rRNA gene, metagenomic and metatranscriptomic sequencing

16S rRNA gene libraries containing the V4 region were prepared from DNA extracted from 24 *C. orbiculata* gill tissues and one *C. orbiculata* foot tissue using protocols described in Chapter II. Metagenomic libraries were prepared from DNA extracted from eight *C. orbiculata* gill samples using NEBNext® dsDNA Fragmentase+NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® or NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA). RNA extracted from eleven *C. orbiculata* gill samples was prepared for metatranscriptomic sequencing using procedures in Chapter II. All library concentrations were quantified with the Qubit® dsDNA HS assay (Life Technologies, Austin, TX, USA) and their insert sizes determined with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). 16S rRNA gene libraries and metagenomic libraries were sequenced on Illumina's MiSeq V2 2x250 bp platform (San Diego, CA, USA) at Clemson University (Clemson, SC, USA), while metatranscriptomic libraries were sequenced on Illumina's HiSeq 4000 2x150bp platform at Duke Center for Genomic and Computational Biology (Durham, NC, USA).

16S rRNA gene sequence analyses

16S rRNA gene reads were processed with Mothur v1.40.5 (Schloss *et al.*, 2009) using methods in Chapter II. 16S rRNA gene OTUs were classified using taxonomy references from SILVA v132 (Quast *et al.*, 2013) and the OTU table was subsampled to

6,599 sequences per sample. Diversity analyses were conducted on subsampled Biological Observation Matrix (BIOM) data exported from by Mothur (Schloss *et al.*, 2009) using the phyloseq v1.16.2 (McMurdie and Holmes, 2013) package in R (R Core Team, 2016). For each calculated alpha diversity measure, R was used to perform the Shapiro-Wilk test (Shapiro and Wilk, 1965) to determine data normality; Levene's test (non-normally distributed data), Bartlett test (Bartlett, 1937; normally distributed data) and Fligner-Killeen test (Fligner and Killeen, 1976; normally distributed data) to determine the homogeneity of variances; Wilcoxon-Mann-Whitney test (Wilcoxon, 1945) with Bonferroni correction for multiple-testing of non-normally distributed, homoscedastic data; Kruskal-Wallis test (Kruskal and Wallis, 1952) for comparing population distributions of non-normally distributed, heteroscedastic data; and analysis of variance (ANOVA) test to compare means across categories for normally distributed, homoscedastic data. Microbiome structure across specimens was compared in phyloseq using the Bray-Curtis dissimilarity index (Bray and Curtis, 1957) and visualized with a PCoA plot. The distance matrix was tested for normality using the Shapiro-Wilk test (Shapiro and Wilk, 1965) and the statistical significance of its categorical partitioning was evaluated with the non-parametric adonis test (Anderson, 2001) implemented in R's vegan v2.4.0 package (Oksanen *et al.*, 2016).

Metagenomic data analyses

Reads from all eight metagenomic libraries were trimmed and assembled individually using procedures in Chapter II. Additionally, two metagenomic libraries of

gill specimens dominated by OTU1 and two libraries of gill specimens dominated by OTU2 were co-assembled separately using the same method to obtain better quality MAGs. Read mapping, binning, MAG quality assessment, MAG annotation, and ANI and AAI calculations were performed as detailed in Chapter II. Bacterial replication rates were estimated using the iRep software (Brown *et al.*, 2016) by mapping metagenomic reads to representative MAGs with ≥75% completeness and ≤2% contamination, Bowtie2's (Langmead and Salzberg, 2012) --no-unal --no-mixed --no-discordant --gbar 1000 --end-to-end -k 200 options. Non-chimeric fragments mapped to protein-coding genes were calculated with featureCounts v1.5.2 (Liao *et al.*, 2014) using the -c and -p options.

Metatranscriptomic data analyses

De novo metatranscriptomic assembly, transcript cluster (gene) abundance estimation, count normalization, transcript-to-gene mapping and transcript annotation were performed using Trinity v2.6.6 (Haas *et al.*, 2013), Trinotate v3.1.1 (https://trinotate.github.io/) and web and local blast searches (NCBI Resource Coordinators, 2016), as documented in Chapter II. Differential expression (DE) and functional gene ontology (GO) enrichment analyses were including 22G (OTU1), 22B+4D (OTU2), 21D (OTU3), and 4F (OTU4), using Bowtie2 v2.3.4.1's very sensitive local and dovetail mode and SAMtools v1.7 (Li *et al.*, 2009). Pan-genomes for each gammaproteobacterial strain and species were created using the method in Chapter II. For symbiont abundance estimation, trimmed reads were mapped to pan-genomes of each symbiont strain and species using performed on raw read counts processed with Trinity's

remove_batch_effects_from_count_matrix.pl script. Batch-removed read counts of transcript clusters (genes) mapped to the thioautotrophic symbiont MAGs and those with homologs in the superkingdom Eukaryota were analyzed separately using DESeq2 (Love *et al.*, 2014), edgeR (Robinson *et al.*, 2010), Reproducibility-Optimized Test Statistic (ROTS; Suomi *et al.*, 2017), voom (Law *et al.*, 2014) and GOSeq (Young *et al.*, 2010) software incorporated within Trinity using a threshold of >2 fold-change and <0.05 false discovery rate (FDR)-adjusted p-value. Predictions made by DE software were compared using Venny (Oliveros, 2007).

Phylogenetic analyses

Phylogenetic analyses of 16S rRNA gene sequences (K2+G model), methanol dehydrogenase (Mdh; LG+G+I+F model) and formate dehydrogenase alpha subunit (FdhA; LG+G+I model) protein sequences were performed using procedures in Chapter II. Phylogenomic analysis (Seah, 2014) was also conducted using methods in Chapter II, and a concatenated alignment of eight single-copy genes (*dna*G, *nus*A, *pgk*, *rplS*, rpsE, *rpsK*, *rps*M, *smpB*) were used to generate a final maximum likelihood tree with aLRT (approximate likelihood-ratio test) SH-like support values (Anisimova and Gascuel, 2006) using the protein substitution model LGF for *dnaG*, *nusA*, *pgk*, *rplS*, *rpsM*, RTREVF for *rpsE* and *rpsK*, and Dayhoff for rpsM and LG for *smpB*.

Host phylogeny was inferred from lucinid marker gene sequences identified in unbinned *C. orbiculata* gill metagenomes using BLAST v2.6.0+'s (NCBI Resource Coordinators, 2016) blastn (Altschul *et al.*, 1990) function and reference sequences

retrieved from GenBank (Benson *et al.*, 2014) via keyword searches. Each identified marker gene set was aligned with BioEdit v7.25's (Hall, 1999) ClustalW (Thompson *et al.*, 1994) package. The cytochrome b gene alignment was analyzed with MEGA7 (Kumar *et al.*, 2016) using the invertebrate mitochondrial genetic code table and the highest scoring Hasegawa-Kishino-Yano model (Hasegawa *et al.*, 1985) with discrete Gamma distribution modeling of the evolutionary rate differences among sites (5 categories (+G, parameter = 0.8357)). All positions of the gap-free alignment were used for phylogenetic analysis and a maximum likelihood (ML) tree with 1,000 bootstrap replicates was generated. 18S rRNA gene and 28S rRNA gene alignments were concatenated and analyzed with RAxML v7.7.2 (Stamatakis, 2006). Twenty runs of initial tree finding were performed with the GTRGAMMA algorithm and the resulting tree was used for the optimization of each nucleotide model and branch lengths. The GTRCATI model yielded a tree with the highest gamma-based likelihood and was used in the final search for the highest-scoring ML tree from 1,000 bootstrap replicates.

qPCR

Primers targeting the *mdh* genes annotated in OTU1-related MAGs (18F and 694R for cloning; 599F and 694R for qPCR) and OTU2-related MAGs (699F and 1159R for cloning; 699F and 804R for qPCR) were designed using Primer3 (Untergasser *et al.*, 2012) in Geneious v8.0 (Kearse *et al.*, 2012; Table 3.1). qPCR standards were prepared from PCR-cloned *mdh* genes as detailed in Chapter II. cDNA concentrations for qPCR were quantified fluorometrically with the Qubit® ssDNA assay (Life Technologies, Austin, TX,

USA). All PCR and qPCR reactions were run on Bio-Rad's C1000 Touch™ Thermal Cycler (Hercules, CA, USA) under the following conditions: Initial denaturation at 95°C for 3 minutes, 29 (cloning)/34 (qPCR) cycles of denaturation at 95°C for 15 seconds, annealing (Table 3.1) for 30 seconds, extension at 72°C for 30 seconds, followed by elongation at 72°C for 5 minutes. qPCR data was analyzed with the CFX Manager software (Bio-Rad Laboratories) and all copy numbers were normalized to the amount (ng) of input DNA/cDNA.

Table 3.1. List of PCR and qPCR primers used in this chapter.

Primer	**Annealing temperature**	**Sequence (5'->3')**
mdh OTU1 18F	57.8°C	TACCCTGCTCGATCCCAAGA
mdh OTU1 599F	59.5°C	CATCCTACTCGCCACGTACC
mdh OTU1 694R	59.5°C (PCR for cloning)/57.8°C (qPCR)	GTTGACCCGCGGTATAGGAG
mdh OTU2 699F	59.5°C (PCR for cloning)/55.2°C (qPCR)	GAAGACCACCCATCTTGGCA
mdh OTU2 804R	59.5°C	GGAGTACCAACCCCAAGTGG
mdh OTU2 1159R	55.2°C	GGCCTGTCTTCATGTCCACA

Availability of data and materials

All specimens are cataloged at the South Dakota School of Mines and Technology, Museum of Geology, with details provided through the iDigBio portal (https://www.idigbio.org/portal/recordsets/db3181c9-48dd-489f-96ab-a5888f5a938c). Sequence data are uploaded to NCBI (NCBI Resource Coordinators, 2016) under the BioProject ID PRJNA377790. Accession numbers are listed in Table 3.2.

Table 3.2. NCBI accession numbers of raw read and sequence data generated in this chapter. All data are linked to NCBI's BioProject ID PRJNA377790 (NCBI Resource Coordinators, 2016).

Database	Accession numbers	Dataset description
Sequence Read Archive (SRA)	SRR5873713-SRR5873738; SRR7235714; SRR7235722-SRR7235725; SRR7235728-SRR7235730	Amplicon-sequenced read data (V4 region of 16S rRNA gene)
	SRR5872870-SRR5872873; SRS3349532-SRS3349535	Metagenomic read data
	SRR7235715-SRR7235721; SRR7235726-SRR7235727; SRR7235731	Metatranscriptomic read data
GenBank	KY687497-KY687506	Sequences of top ten most abundant OTUs
	NATR00000000-NATW00000000; QBVC00000000-QBVG00000000	Metagenome-assembled genomes (MAGs)

Results

Site characterization

Ctena orbiculata specimens were collected from Sammy Creek Landing, Sugarloaf Key, Florida in June 2016 within quadrats predominantly covered with either seagrass (*Halodule* sp. and *Thalassia* sp.) or algae (*Halimeda* sp. and *Penicillus* sp.; Figure 3.1). Sampled quadrats comprised of 84% live *C. orbiculata* specimens, 9% live *Lucinisca nassula* specimens, 5% live *Anodontia alba* specimens, and 2% live *Codakia orbicularis* specimens (Table 3.3). Small numbers (<5) of live *Parvilucina pectinella* and *Radiolucina amianta* were also collected at the site, but outside the quadrats used in this study. Estimated densities of live *C. orbiculata* clams per m^3 were higher in the sampled algae-covered quadrats (117±33) than seagrass-covered quadrats (23±3; Table 3.3). No clear geochemical differences were observed between sampled seagrass-covered and algae-covered quadrats. In these quadrats, porewater and ocean water temperatures (30-31°C) and pH (7-8) were stable (Table 3.3). Porewater dissolved oxygen concentrations in these quadrats ranged from 0.2 to 0.8 mg/L and were lower than corresponding oxygen levels in the ocean water (3 to 6 mg/L); Table 3.3). Porewater sulfide and methane concentrations in these quadrats varied between 2 to 20 mg/L and 0.3 to 22 μg/L, respectively (Table 3.3).

Table 3.3. Environmental data from Sammy Creek Landing, Sugarloaf Key, Florida. Quadrats are numbered as transect number, then distance from the shoreline (in meters). The estimated density of live clams was calculated from the number of live clams recovered from the volume of sediment excavated per quadrat. Quadrats adjacent to (T22 0m) or where *C. orbiculata* specimens were sequenced in this chapter were highlighted in grey.

Quadrat	Vol (m^3)	Seagrass or algae species (%)	Estimated density of live clam species per m^3						Pore water[1]					Ocean water[1]			
			Ctena	*Lucinisca*	*Anodontia*	*Codakia*	*Parvilucina*	*Radiolucina*	Sulfide	CH_4	DO	Temp	pH	DO	Temp	pH	Cond
T20 (0m)	0.42	*Halodule* (50); *Syringodium* (50)	14.1	2.5	0	0	0	0	98.83	34.94	0.12	31.0	7.21	3.36	30.2	7.90	69.2
T20 (10m)	0.40	*Halodule* (100)	2.5	0	0	0	0	0	55.88	6.03	0.54	30.2	7.24	5.40	31.2	8.15	69.7
T20 (20m)	0.29	*Halodule* (100)	20.7	3.5	0	0	0	0	20.35	4.29	0.51	31.1	7.27	5.60	31.4	8.29	69.5
T20 (30m)	0.30	*Thalassia* (60); Mixed red algae (30); *Halimeda* algae (10)	16.7	0	3.3	0	0	0	3.53	0.38	0.88	29.3	7.73	5.49	29.3	8.02	65.3
T20 (40m)	0.40	*Halodule* (90); *Thalassia* (10)	25	2.5	2.5	0	0	0	5.49	0.32	0.28	30.0	7.93	5.56	30.5	8.22	69.0
T20 (50m)	0.43	*Halodule* (30); *Thalassia* (70); trace *Halomeda* algae	18.6	2.3	7	0	0	0	18.83	0.56	1.20	31.1	7.67	6.59	31.9	8.14	69.6
T21 (0m)	0.07	*Halimeda* algae (70); *Pennicilus* (30); trace *Halodule*	93.3	0	0	0	0	0	20.00	22.41	0.69	30.5	7.44	2.85	30.8	7.85	67.2
T21 (10m)	0.10	*Halodule* (100); trace *Pennicilus*	80	10	0	30	0	0	4.53	0.51	0.79	31.2	7.69	4.63	31.3	7.96	68.5
T21 (20m)	0.10	*Halodule* (100)	20	20	0	130	20	0	0.32	0.31	1.25	31.4	7.68	4.55	32.1	8.04	69.9
T21 (30m)	0.10	*Halodule* (100); trace *Thalassia*	60	30	10	60	0	10	12.17	0.25	0.59	32.0	7.52	4.95	32.4	7.96	70.9
T21 (40m)	0.10	*Halodule* (90); mixed green algae (10)	40	0	20	10	0	0	3.95	0.34	0.92	32.1	7.6	5.66	33.4	8.17	72.3
T21 (50m)	0.10	*Halodule* (100); trace *Halimeda*	40	10	0	50	0	0	11.20	0.19	0.47	31.7	7.7	4.11	31.6	7.93	69.0
T22 (0m)	0.10	*Halimeda* (100); trace *Halodule*	140	0	10	10	0	0	2.01	0.61	0.80	32.0	7.51	5.22	32.3	7.76	69.7

[1] Dissolved sulfide and dissolved oxygen (DO) measurements are reported in mg/L. Dissolved methane concentrations are reported in µg/L. Temperature is in Celsius. Conductivity (Cond) is in mS/cm.

Gill microbiome diversity

Sequenced 16S rRNA genes (V4 region) from 24 *C. orbiculata* gill tissues and one *C. orbiculata* foot tissue were clustered into Operational Taxonomic Units (OTUs) at 99% identity for higher species resolution (Edgar, 2018) and showed 99±0.02% average Good's sequencing coverage (Good, 1953; Figure 3.2a). Five co-existing *Ca.* Thiodiazotropha-like OTUs (OTUs 1-5) were present at >60% relative abundances in at least one gill specimen of the sampled host population (Figure 3.2a). OTU1 dominated 17 of 24 of gill specimens with average 94±5% relative abundance, while OTU2 dominated three specimens (82±5% average relative abundance) and OTU3 dominated two specimens (97±1% average relative abundance; Figure 3.2a). OTUs 4 and 5 dominated one specimen each at 96% and 69% relative abundances, respectively (Figure 3.2a). OTU4 was also identified in the OTU1-dominated gill specimen 4E at 0.2% relative abundance (Figure 3.2a). The OTU5-dominated gill specimen 2B had the lowest Good's coverage (Good, 1953) of 88% and highest alpha diversity values compared to other specimens, while OTU2-dominated gill specimens were predicted to have significantly higher Shannon (Shannon, 1948; Wilcoxon-Mann-Whitney p=0.0053) and Simpson indices (Simpson, 1949; Wilcoxon-Mann-Whitney p=0.021) than OTU1-dominated gill specimens (Figures 3.2-3.3). Gill microbiome structures dominated by different OTUs were statistically different from one another (adonis R=0.98, p=0.001; Figure 3.2c). The OTU distribution did not follow any clear spatial trend; high abundances of OTUs 1-3 were identified in gill specimens in both seagrass- and algae-covered quadrats (Figures 3.2a). Two OTU4-dominated gill specimens, however, occurred in the same seagrass-covered quadrat (Figure 3.2a).

In 69% of the sampled gill microbiomes, we also identified 2±3% relative abundances of a gammaproteobacterial *Endozoicomonas*-like OTU (OTU9; order Oceanospirillales) most closely related to *E. elysicola* from the gastrointestinal tract of the sea slug *Elysia ornata* (Kurahashi and Yokota, 2007; Figure 3.2a and Figure 3.4). We could not assemble the genome of the *Endocoizomonas*-like (OTU9; order Oceanospirillales) species from the gill metagenomes, possibly because of the lack of sequencing coverage on the MiSeq platform. The presence of *Endocoizomonas*-like OTU9 in the gill microbiomes corroborates previous reports of potentially taxonomically and/or functionally-related bacterial members enriched in lucinid gill microbiomes, including other gammaproteobacterial taxa in *Loripes orbiculatus* (Mausz *et al.*, 2010 unpublished - NCBI accession numbers GQ853555- GQ853555 and Espinosa *et al.*, 2013), *Kistimonas*-like species (order Oceanospirillales) in *Phacoides pectinatus* (Chapter II), and an unclassified rod-shaped taxon in *Euanodontia ovum* (Ball *et al.*, 2009). *Endozoicomonas* species have also been identified as coral symbionts (Neave *et al.*, 2014; Neave *et al.*, 2017; Neave *et al.*, 2016; Ding *et al.*, 2016), sea squirt commensals (Schreiber *et al.*, 2016) and a cobia fish pathogen (Mendoza *et al.*, 2013). The single sequenced foot specimen was dominated by a *Spirochaeta*-like OTU (OTU6; 90% relative abundance) related to the spirochete symbiont in *Lucinoma aff. kazani* (Duperron *et al.*, 2012; Figure 3.2a and Figure 3.4). The foot is crucial to porewater sulfide uptake (Taylor and Glover, 2000; Taylor and Glover, 2010), oxygenated water and nutrient transport (Taylor and Glover, 2000; Taylor and Glover, 2010) and perhaps symbiont acquisition in lucinid clams, but its microbiome has not been comprehensively investigated. Spirochetes have been reported in the gills of

P. pectinatus (Chapter II), in gutless oligochete worms as symbionts (Blazejak *et al.*, 2005; Ruehland *et al.*, 2008), and in the epibiont of a hydrothermal vent worm epibiont (Campbell and Cary, 2001). Currently, the roles of Oceanospirillales and spirochete species in the lucinid gill and foot microbiomes remain unknown.

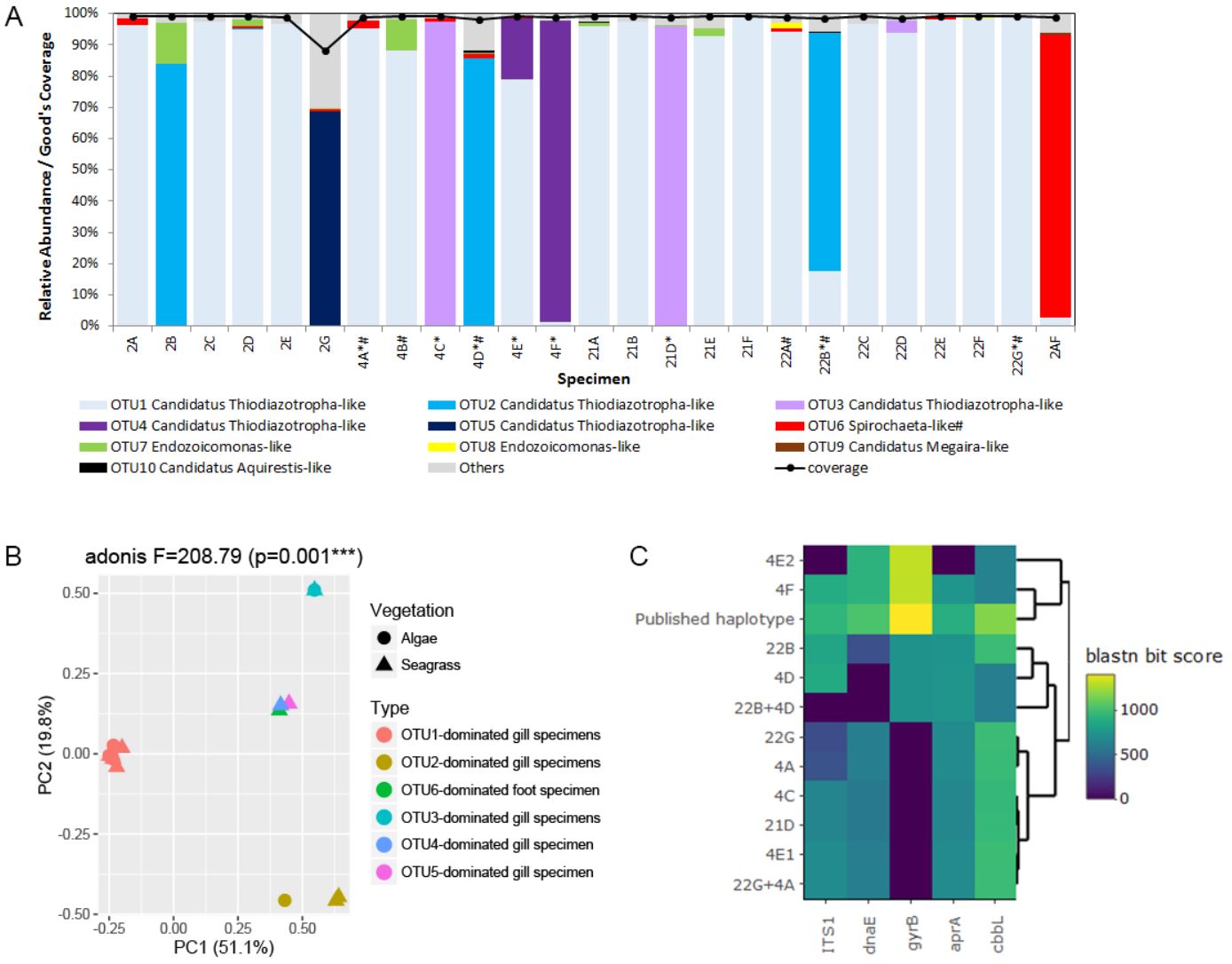


Figure 3.2. (A) Relative abundances of subsampled bacterial OTUs identified in *C. orbiculata* gill and foot (specimen 2AF) specimens. '*' and '#' denote specimens also used for metagenomic and metatranscriptomic sequencing, respectively. (B) PCoA plot showing differences in microbiome community structure (Bray-Curtis index) among gill specimens dominated by different OTUs. (C) blastn (Altschul *et al.*, 1990) bit scores of pairwise sequence comparisons between published marker gene haplotype sequences from the Antillean *C. orbiculata* symbiont (Brissac *et al.*, 2016) and corresponding sequences of gammaproteobacterial MAGs identified in this study. Abbreviations: ITS1, 16S-23S rRNA intergenic spacer region; *dnaE*, DNA polymerase III alpha subunit; *gyrB*, DNA gyrase B subunit; *aprA*, adenylylsulfate reductase alpha subunit; *cbbL,* ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit.

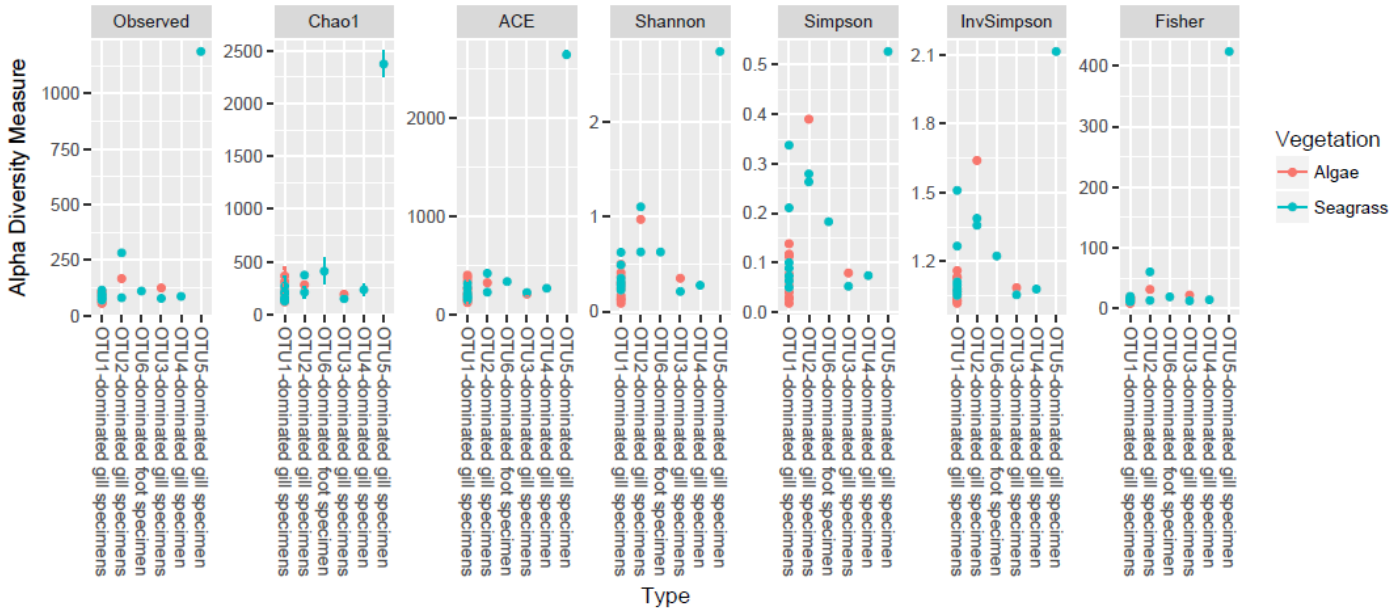


Figure 3.3. Alpha diversity measures calculated for each amplicon-sequenced gill specimen.


16S rRNA gene
bootstrap consensus tree
Myrtea tanimbarensis gill symbiont [FN869533]
Lucinidae sp. gill symbiont [FN869549]
Gloverina rectangularis gill symbiont [FN869545]
Codakia orbicularis gill symbiont (Bahamas) [M99447]
Ca. Thiodiazotropha endolucinida (Carribean) [X84979/MARB00000000]
Ca. Thiodiazotropha-like OTU4 [KY687501]
Ca. Thiodiazotropha-like MAG 4F [QBVD01000000]
Lucina nassula gill symbiont [X95229]
Ca. Thiodiazotropha-like MAG 21D [QBVC00000000]
Ca. Thiodiazotropha-like OTU3 [KY687500]
Ca. Thiodiazotropha-like MAG 4C [QBVG00000000]
Ca. Thiodiazotropha-like OTU2 [KY687498]
Ca. Thiodiazotropha-like MAG 22B+4D [NATW00000000]
Myrtea flabelliformis gill symbiont [FN869544]
Cardiolucina quadrata gill symbiont [FN869540]
Lucinoma aff. kazani gill symbiont [AM236336]
Lucinoma annulata gill symbiont [M99449]
Lucinoma aequizonata gill symbiont [M99448]
Dulcina sp. gill symbiont [FN869536]
Ca. Thiodiazotropha endoloripes [LVJZ01000003]
Codakia costata gill symbiont [L25712]
Stewartia floridana gill symbiont [L25707]
Ca. Thiodiazotropha-like OTU5 [KY687502]
Ca. Thiodiazotropha-like MAG 4E1 [QBVF00000000]
Ca. Thiodiazotropha-like OTU1 [KY687497]
Ca. Thiodiazotropha-like MAG 22G+4A [NATV00000000]
Ca. Thiodiazotropha-like MAG 22G [NATS00000000]
Ca. Thiodiazotropha-like MAG 4A [NATT00000000]
Clade A
Ca. Thiodiazotropha spp.
Bathymodiolus thermophilus gill symbiont [M99445]
Bathymodiolus septemdierum gill symbiont [AB598130]
Uncultured SUP05 cluster bacterium clone BFB022 [KC545709]
Calyptogena magnifica gill symbiont [M99446]
Calyptogena pacifica gill symbiont [AF035723]
Uncultured SUP05 cluster bacterium clone J8P41000 2H04 [GQ351132]
Sedimenticola selenatireducens [NR_041877]
Sedimenticola thiotaurini [NR_136790]
Thyasira flexuosa gill symbiont [L01575]
Kuphus polythalamia gill symbiont [NZ_CP019936]
Anodontia phillipiana gill symbiont [L25711]
Anodontia ovum gill symbiont [EU98357]
Clade B
Riftia pachyptila trophosome symbiont [M99451]
Ridgeia piscesae trophosome symbiont [JX570594]
Oasisa sp. trophosome symbiont [EF675613]
Escarpia spicata trophosome symbiont [AF165909]
Phacoides pectinatus gill symbiont [X84980]
Ca. Sedimenticola endophacoides [KY509297]
Clade C
Thiomicrospira sp. [L01576]
Solemya velum gill symbiont [M90415]
Allochromatium vinosum [NR_044605]
Thiocapsa marina [AJ581994]
Marichromatium purpuratum [KP212379]
Olavius algarvensis gammaproteobacterial endosymbiont [KP94394]
Inanidrilus leukodermatus endosymbiont [U24110]
Laxus sp. symbiont [U14727]
Ca. Thiosymbion oneisti [KT826595]
Thiothrix nivea [L40993]
Teredinibacter turnerae [NC_012997]
Osedax japonicus heterotrophic symbiont R46 [AB293969]
Neptunomonas japonica [AB291228]
Osedax japonicus heterotrophic symbiont R21 [AB293970]
Osedax sp. Rs2 symbiont [KC862279]
Osedax sp. clone Rs1 symbiont [KC862278]
Kistimonas scapharcae [NR_109422]
Phacoides pectinatus Kistimonas-like OTU2 [KY509298]
Phacoides pectinatus Kistimonas-like OTU6 [KY509302]
Kistimonas asteriae [NR_116386]
Acesta excavata intracellular Oceanospirilalles clone Ae1pa1 [EF508132]
Kistimonas sp. bin40 [MPMQ01000002]
Ca. Endonucleobacter bathymodioli [FM162182]
Loripes orbiculatus clone1B [GQ853556]
Loripes orbiculatus clone1A [GQ853555]
Endozoicomonas montiporae [KJ372452]
Endozoicomonas elysicola [EU919205]
Endozoicomonas-like OTU7 [KY687503]
Endozoicomonas-like OTU8 [KY687504]
Ca. Megaira-like OTU9 [KY687505]
Olavius algarvensis spirochete endosymbiont [AJ620502]
Olavius loisae spirochete endosymbiont [AF104475]
Phacoides pectinatus Spirochaeta-like OTU5 [KY509301]
Lucinid sp. spirochete endosymbiont [AM236337]
Spirochaeta-like OTU3 [KY687499]
Termite gut Spirochaeta sp. [AB015819]
Spirochaeta halophila [NR_044756]
Spirochaeta cellobiosiphila [NR_044505]
Ca. Aquirestis-like OTU10 [KY687506]
Desulfurobacterium thermolithotrophum [NR _075040]

Figure 3.4. Maximum likelihood tree of the ten most abundant 16S rRNA gene OTUs and 16S rRNA gene sequences recovered from metagenome-assembled genomes (MAGs) discovered in this study (red text), in relation to symbionts of other lucinid species (blue text), marine species, insect species and free-living bacterial species. GenBank (Benson *et al.*, 2014) accession numbers are indicated in square brackets and bootstrap values of >70% are displayed. The outgroup used was *Desulfurobacterium thermolithotrophum* from phylum Aquificae.

Metagenomic sequencing of a subset of eight gill specimens dominated by *Ca.* Thiodiazotropha-like OTUs 1-4 yielded four OTU-specific clusters of gammaproteobacterial metagenome-assembled genomes (MAGs; Figure 3.5 and Table 3.4). 16S rRNA gene sequences annotated in OTUs 1-4-related MAGs shared 99-100% identity in their V4 regions with their corresponding OTU sequences from the same gill specimens. Phylogenetic analyses using the 16S rRNA gene and eight marker genes grouped these MAGs with other clade A thioautotrophic lucinid endosymbionts and OTU-specific clustering was observed on both trees (Figure 3.4-3.5). 16S rRNA gene and five other marker gene sequences of the OTU4-related MAG were most similar to haplotype sequences identified from the thioautotrophic *C. orbiculata* symbiont at Lesser Antilles, French West Indies (Brissac *et al.*, 2016; Figure 3.2d).

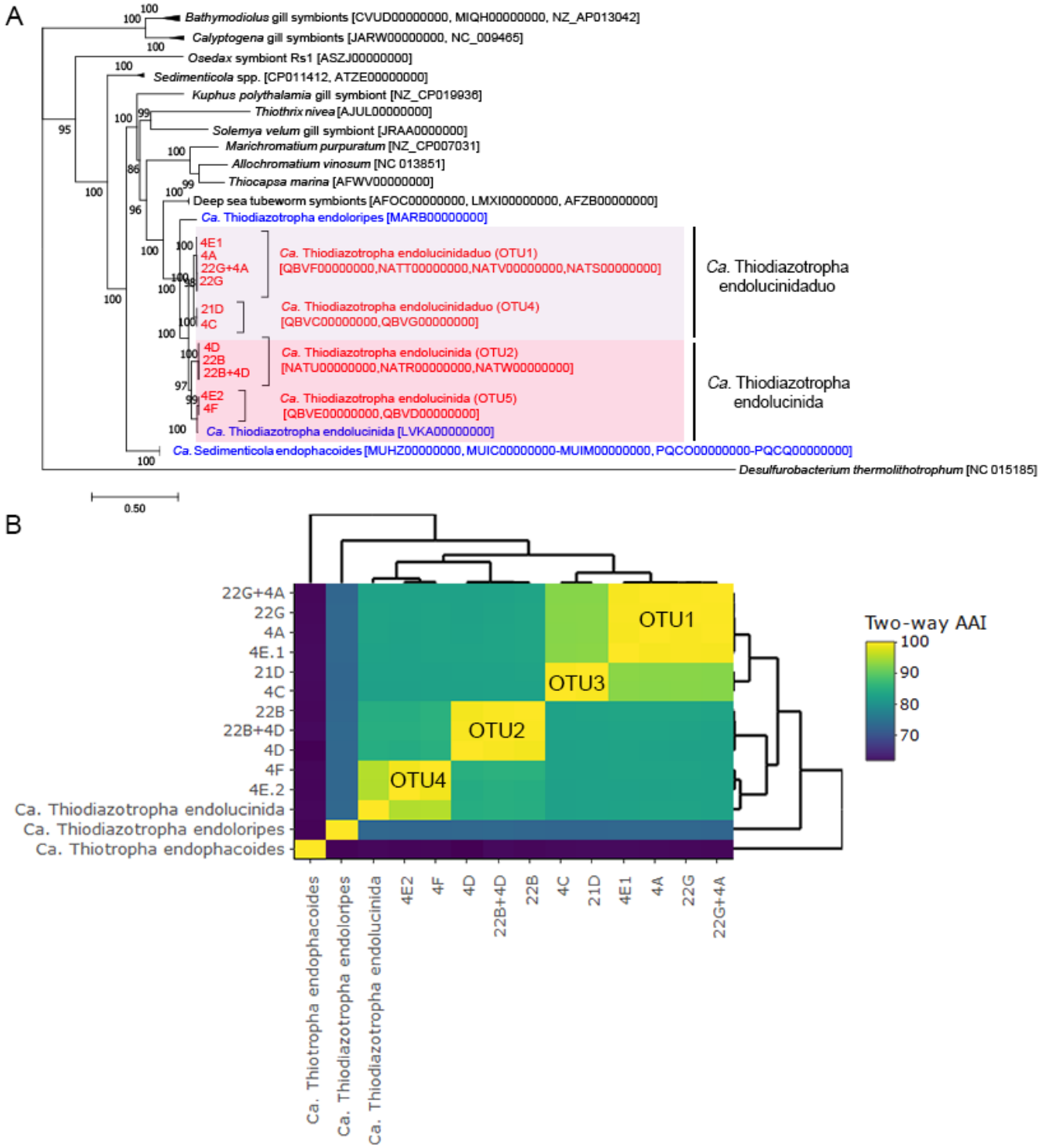

A
Bathymodiolus gill symbionts [CVUD00000000, MIQH00000000, NZ_AP013042]
Calyptogena gill symbionts [JARW00000000, NC_009465]
Osedax symbiont Rs1 [ASZJ00000000]
Sedimenticola spp. [CP011412, ATZE00000000]
Kuphus polythalamia gill symbiont [NZ_CP019936]
Thiothrix nivea [AJUL00000000]
Solemya velum gill symbiont [JRAA00000000]
Marichromatium purpuratum [NZ_CP007031]
Allochromatium vinosum [NC 013851]
Thiocapsa marina [AFWV00000000]
Deep sea tubeworm symbionts [AFOC00000000, LMXI00000000, AFZB00000000]
Ca. Thiodiazotropha endoloripes [MARB00000000]
4E1
4A
22G+4A
22G
Ca. Thiodiazotropha endolucinidaduo (OTU1)
[QBVF00000000,NATT00000000,NATV00000000,NATS00000000]
21D
4C
Ca. Thiodiazotropha endolucinidaduo (OTU4)
[QBVC00000000,QBVG00000000]
Ca. Thiodiazotropha endolucinidaduo
4D
22B
22B+4D
Ca. Thiodiazotropha endolucinida (OTU2)
[NATU00000000,NATR00000000,NATW00000000]
4E2
4F
Ca. Thiodiazotropha endolucinida (OTU5)
[QBVE00000000,QBVD00000000]
Ca. Thiodiazotropha endolucinida
Ca. Thiodiazotropha endolucinida [LVKA00000000]
Ca. Sedimenticola endophacoides [MUHZ00000000, MUIC00000000-MUIM00000000, PQCO00000000-PQCQ00000000]
Desulfurobacterium thermolithotrophum [NC 015185]
0.50
B
22G+4A
22G
4A
4E.1
21D
4C
22B
22B+4D
4D
4F
4E.2
Ca. Thiodiazotropha endolucinida
Ca. Thiodiazotropha endoloripes
Ca. Thiotropha endophacoides
OTU1
OTU3
OTU2
OTU4
Two-way AAI
100
90
80
70
Ca. Thiotropha endophacoides
Ca. Thiodiazotropha endoloripes
Ca. Thiodiazotropha endolucinida
4E2
4F
4D
22B+4D
22B
4C
21D
4E1
4A
22G
22G+4A

Figure 3.5. (A) Phylogenomic tree of gammaproteobacterial MAGs identified in this study in relation to lucinid (blue), bivalve, tubeworm symbionts and free-living bacteria, based on eight single-copy marker genes (*dnaG*, *nusA*, *pgk*, *rplS*, *rpsE*, *rpsK*, *rpsM*, *smpB*). GenBank (Benson *et al.*, 2014) accession numbers of other sequences are indicated in square brackets. The outgroup used was *Desulfurobacterium thermolithotrophum* from phylum Aquificae. Tree nodes show approximate likelihood-ratio test (aLRT) SH-like support values (Anisimova and Gascuel, 2006). The scale bar indicates 0.5 substitution per site. (B) Heatmap of pairwise AAI comparisons across gammaproteobacterial symbiont MAGs identified in this study and other lucinid species.

MAGs within each of the four OTU-specific clusters shared >99% pairwise average nucleotide identity (pANI) and average amino identity (pAAI) and were most closely related to the representative MAG of *Ca.* Thiodiazotropha endolucinida (König *et al.*, 2016; Figure 3.5b). OTU2-related MAGs shared 75±1% pANI and 85±0.2% pAAI with the *Ca.* Thiodiazotropha endolucinida MAG, and 81±1% pANI and 85±0.4% pAAI with OTU4-related MAGs (Figure 3.5b). Similarly, OTU4-related MAGs shared 73±1% pANI and 95±0.08% pAAI with the *Ca.* Thiodiazotropha endolucinida MAG (Figure 3.5b). Species classification of the MAGs, based on the 93-95% pANI and 85-90% pAAI species delineation proposed in Rodriguez-R and Konstantinidis (2014), suggested OTU2, OTU4 and *Ca.* Thiodiazotropha endolucinida to be the same species and we propose the use of the same species name, *Ca.* Thiodiazotropha endolucinida, for OTU2 and OTU4. Because OTU2, OTU4 and *Ca.* Thiodiazotropha endolucinida shared <90% pANI and pAAI with

each other, we posit that they represent different strains of the same species. Based on observed pANI and pAAI comparisons, OTU1 and OTU3 likely represent a species separate from *Ca.* Thiodiazotropha endolucinida (71±5% pANI; 83±0.1% pAAI), OTU2 (80±0.7% pANI; 83±0.2% pAAI) and OTU4 (79±0.8% pANI; 83±0.2% pAAI; Figure 3.5b). Based on the 55-60% pAAI genus boundary proposed in Rodriguez-R and Konstantinidis (2014) and their thiodiazotrophic functional potential (discussed below), OTUs 1-4 and *Ca.* Thiodiazotropha endolucinida can be plausibly classified under the same genus. As such, we propose a new species name within the same genus, *Ca.* Thiodiazotropha endolucinidaduo, for OTU1 and OTU3, where suffix "duo" means "two" in Latin. OTU1-related MAGs and OTU3-related MAGs, which shared 91±0.2% pANI and 93±0.06% pAAI to each other, were likely different strains of the same species (Figure 3.5b).

Metagenomic read coverage profiles of each representative OTU-specific MAG, bacterial replication rates estimated by iRep (Brown *et al.*, 2016) from MAG data, and percentages of metatranscriptomic reads mapped to protein-coding genes of each representative OTU-specific MAG were generally consistent with relative abundance patterns of their corresponding 16S rRNA gene OTUS (Figure 3.6). The only exception, OTU2-dominated gill metatranscriptome 22B, showed higher percentages of reads mapped to OTU1 compared to OTU2 (Figure 3.6c). Symbiont-specific gill transcriptomes of OTU1/OTU3-dominated specimens clustered together with 0.9±0.07 average pairwise Pearson correlation coefficient (PCC), while the OTU2-dominated symbiont transcriptome 4D and OTU4-dominated symbiont transcriptome 4F appeared to be outliers sharing <0.4

PCC with the other specimens (Figure 3.7). Symbiont OTU-specific patterns were not observed across the entire gill metatranscriptomes (average 0.7±0.4 PCC between samples) or another subset of Mollusca-related transcriptomes (average 0.8±0.06 between samples; Figures 3.8-3.9). Host 18S rRNA gene, 28S rRNA gene and mitochondrial cytochrome b (*cytob*) gene sequences extracted from unbinned *C. orbiculata* gill metagenomes clustered unambiguously with reference sequences from *C. orbiculata* (Figure 3.10), confirming the host taxonomy of these specimens.

Table 3.4. Features of metagenome-assembled genomes (MAGs) recovered from *C. orbiculata* gill specimens. Abbreviation: PE, paired-end; PEGs, protein-encoding genes.

Categorized species/strain	MAG ID	# PE Reads	Size	Contigs	PEGs	G+C	N50	Completeness		Contamination	Strain heterogeneity	MAG quality[c]
		M	Mb	#	#	%	Kb	CheckM[a]	BUSCO[b]	%	%	
***Ca.* Thiodiazotropha endolucinida**												
OTU1	22G	0.5	4.0	235	3,718	56	26	97	94	2	0	High
	4A	0.5	3.9	350	3,638	56	16	94	89	2	0	High
	22G+4A	1	4.3	182	3,944	56	36	98	96	2	17	High
OTU3	4E.1	5.8	4.2	122	3,833	56	50	98	96	3	33	High
	4C	6.2	4.6	42	4,070	57	211	98	94	2	20	High
	21D	5.1	4.5	51	4,049	58	139	98	95	2	0	High
***Ca.* Thiodiazotropha endolucinidaduo**												
OTU2	22B	0.3	4.0	309	3,700	54	19	97	88	4	28	Medium
	4D	0.5	3.4	656	3,127	53	6	84	69	2	0	Medium
	22B+4D	0.8	4.1	193	3,729	53	34	98	96	3	8	Medium
OTU4	4E.2	5.8	4	593	3,730	53	9	91	77	4	37	Medium
	4F	7.2	4.7	112	4352	52	90	99	98	2	0	High

[a]Parks *et al.*, 2015; [b]Simao *et al.*, 2015; [c]Bowers *et al.*, 2017

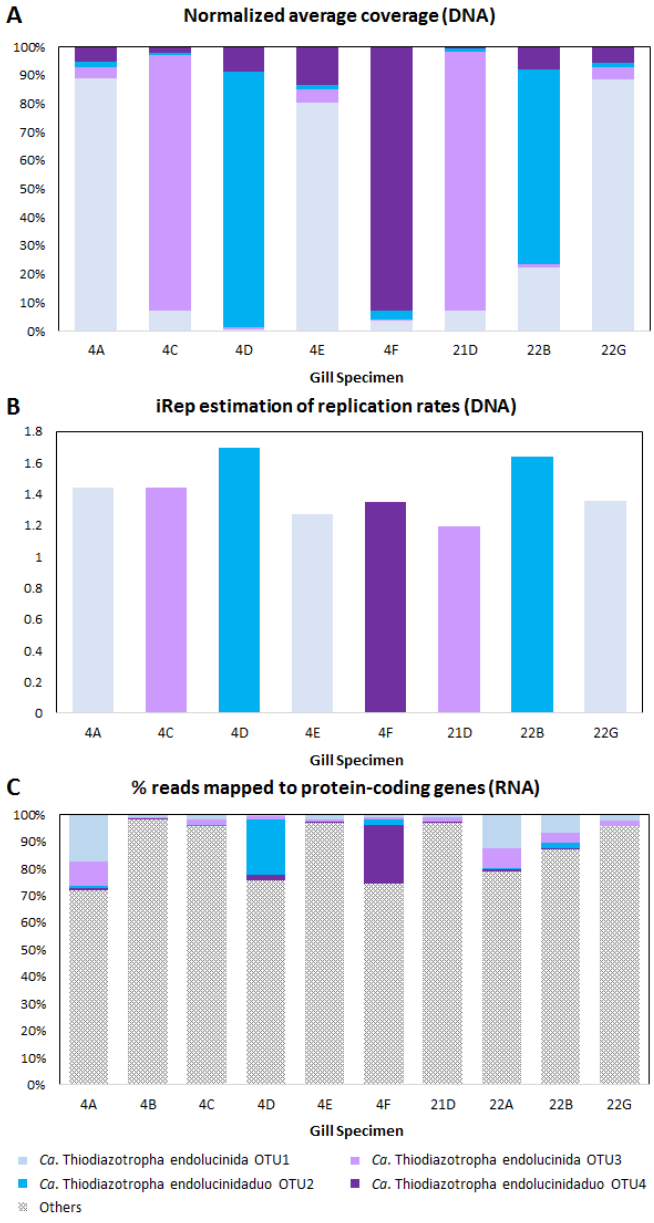


Figure 3.6. (A) Percentage average coverage depths normalized by MAG size (B) iRep (Brown *et al.*, 2016) estimation of replicate rates, and (C) percentage metatranscriptomic reads of each sequenced gill specimen mapped to each representative taxon-specific MAG. Only bars with ≥0 estimated replication rate were shown in (B). Metatranscriptomic reads in (C) were mapped to protein-coding genes of each representative MAG.

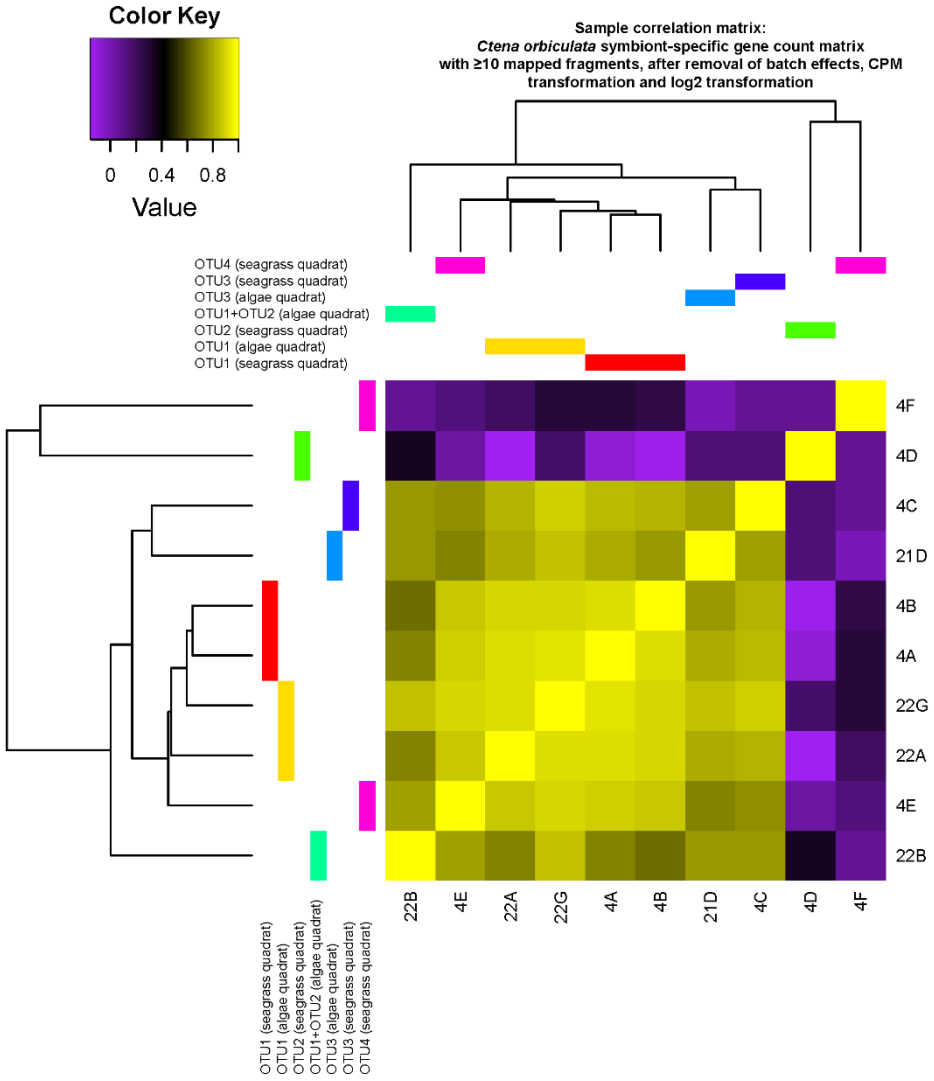


Figure 3.7. Heatmap of pairwise Pearson correlations across gill specimens based on the number of assembled transcripts mapped to genes in symbiont transcriptomes extracted from the metatranscriptomic assembly. The count matrix was processed to filter out genes with <10 mapped fragments, eliminate batch effects, and normalized to $\log_2$ counts per million (CPM).

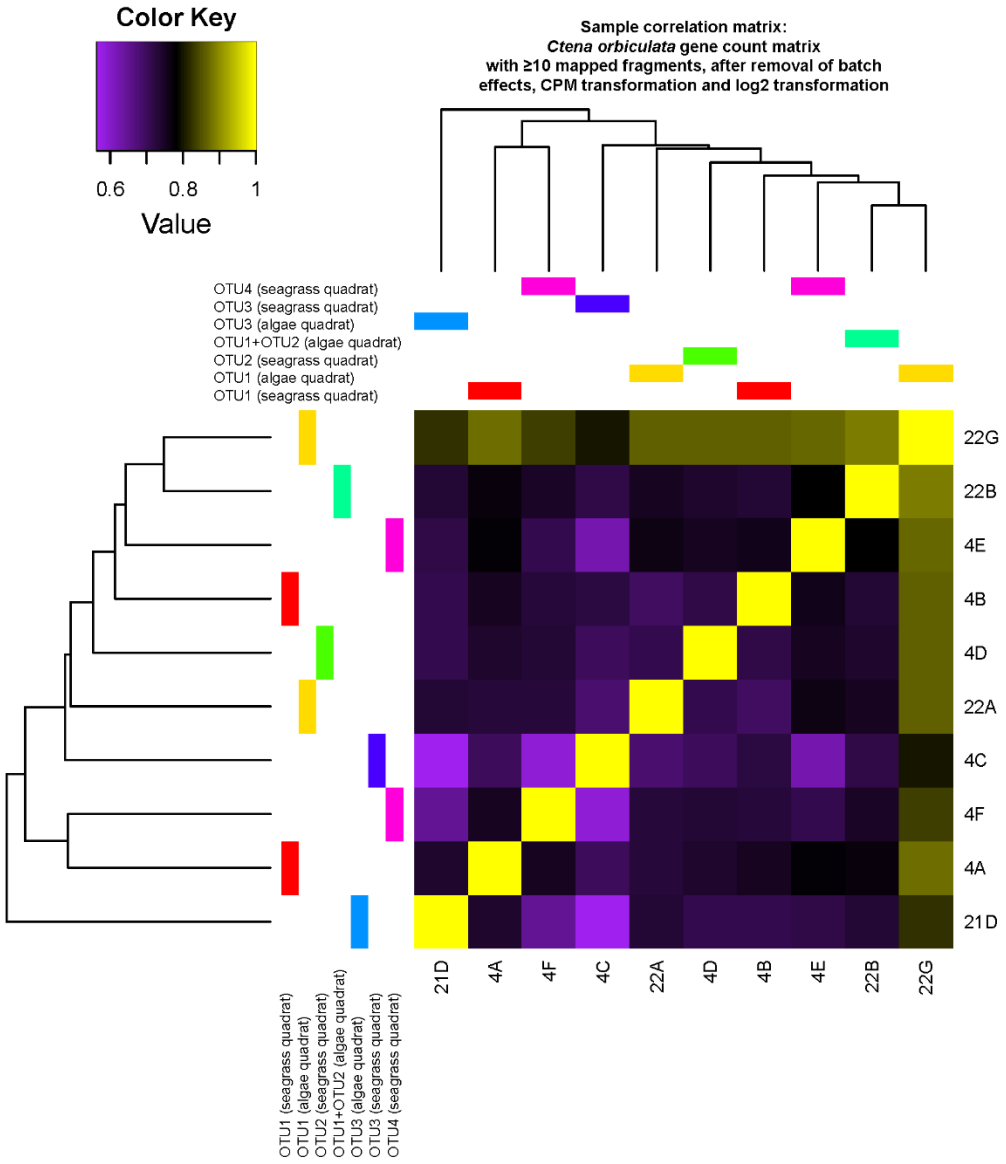


Figure 3.8. Heatmap of pairwise Pearson correlations across gill specimens based on the number of assembled transcripts mapped to genes in the gill metatranscriptomic assembly. The count matrix was processed to filter out genes with <10 mapped fragments, eliminate batch effects and normalized to $\log_2$ counts per million (CPM).

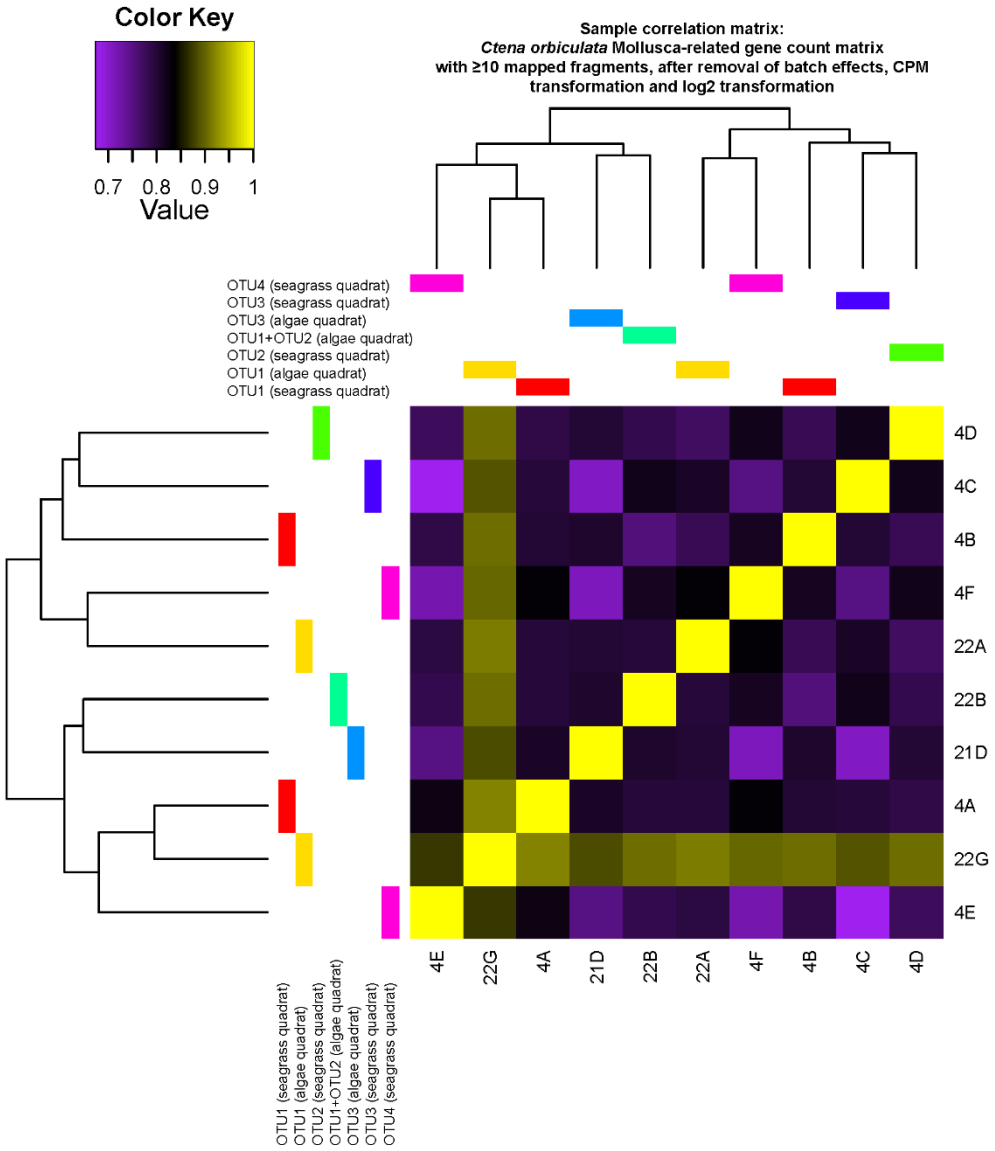


Figure 3.9. Heatmap of pairwise Pearson correlations across gill specimens based on the number of assembled transcripts mapped to Swissprot-annotated (The UniProt Consortium, 2015). Mollusca-related genes in the metatranscriptomic assembly. The count matrix was processed to filter out genes with <10 mapped fragments, eliminate batch effects and normalized to $\log_2$ counts per million (CPM).

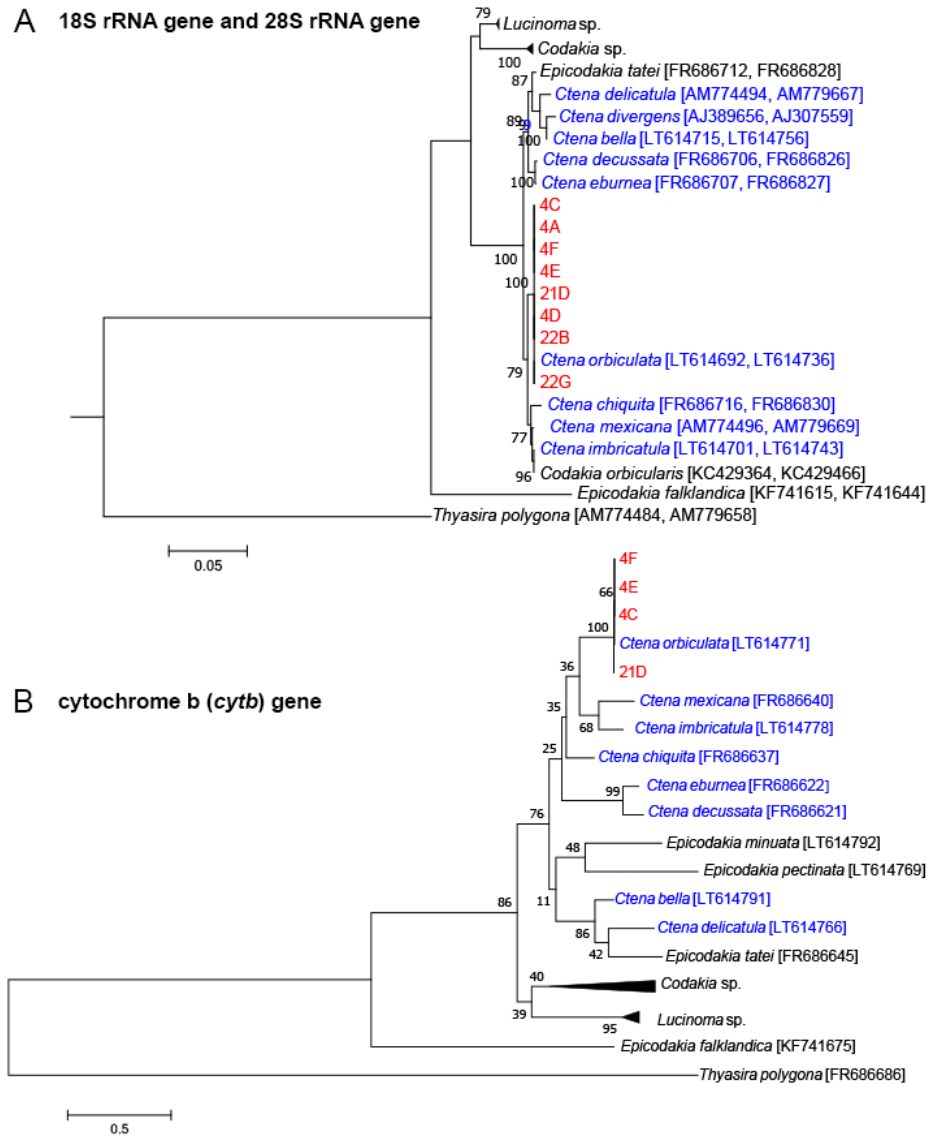


Figure 3.10. Maximum likelihood tree of (A) 18S rRNA gene and 28S rRNA gene sequences and (B) cytochrome b gene sequences from *C. orbiculata* in relation to reference lucinid species. Tree nodes show bootstrap values and square brackets contain GenBank accession numbers for reference sequences. *Thyasira polygona* (order: Lucinida, family: Thyasiridae) was used as the outgroup in both trees. Scale bars indicate the number of substitutions per site.

Core symbiont functions

Pan-genomes of *Ca.* Thiodiazotropha-like MAGs were predicted by Rapid Annotation using Subsystem Technology (RAST; Aziz *et al.*, 2008) to share ~62% gene and ~83% subsystem content (Figure 3.11). It should be noted, as discussed in Chapter II, that gene and subsystem annotations were based on incompletely sequenced and annotated MAGs. As such, the numbers of shared genes and subsystems were imprecise estimates not accounting for missing, unbinned or unclassifiable genes, incomplete pathways and strain/cross-species contamination of the MAGs. Limitations of host-symbiont metatranscriptomic analyses, detailed in Chapter II, also apply to this study.

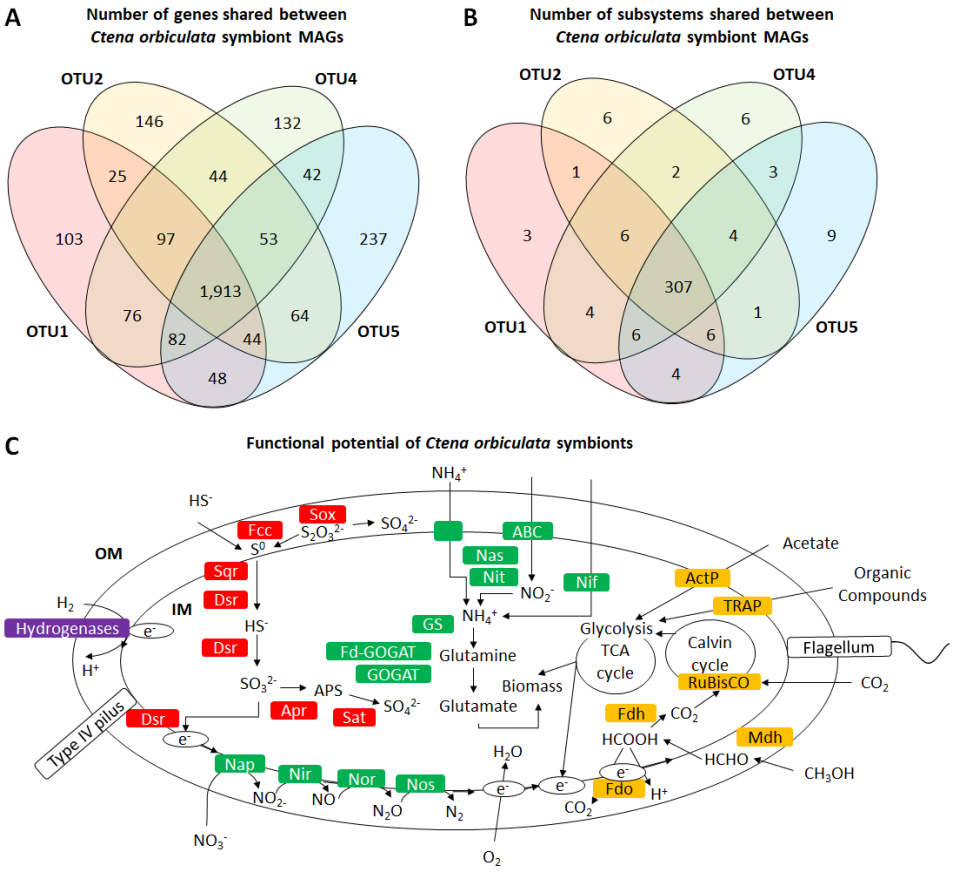


Figure 3.11. Overview of (A-B) RAST-annotated genes and (C) key hydrogen (purple), sulfur (red), nitrogen (green) and carbon (orange) pathways shared among pan-genomes of *C. orbiculata* symbionts. Abbreviations: S^0, elemental sulfur; Fcc, Flavocytochrome c - sulfide dehydrogenase; Sox, sulfur oxidation gene cluster; Sqr, sulfide:quinone oxidoreductases; Dsr, reverse dissimilatory sulfite reductase; Apr, adenylylsulfate reductase; APS, adenosine-5'-phosphosulfate; Sat, sulfate adenylyltransferase; ABC, ATP-binding cassette transporters; GS, glutamine synthetase; Fd-GOGAT, ferrodoxin-dependent glutamate synthase; GOGAT, glutamine oxoglutarate aminotransferase (glutamate synthase); Nas, assimilatory nitrate reductase; Nit, assimilatory nitrite reductase; Nif, nitrogen fixation gene cluster; Nap, periplasmic dissimilatory nitrate reductase; Nir, cytochrome nitrite reductase cd1; Nor, nitric oxide reductase; Nos, nitrous oxide reductase; ActP, acetate permease; TRAP, Tripartite ATP-independent periplasmic transport; RuBisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; IM, inner membrane; OM, outer membrane.

Ctena orbiculata symbionts showed high expression of a carbon storage regulator (103±97 average trimmed mean of M-values normalized transcripts per million; TPM) and thioautotrophy-related form IAq ribulose-1,5 bisphosphate carboxylase/oxygenase (RuBisCO; average 18±19 TPM) and adenylsulfate reductase subunit A (average 15±12 TPM), which were among the 35 most abundant transcript clusters (loosely equivalent to genes) in the symbiont transcriptomes (Figure 3.12). Many stress-related symbiont transcript clusters encoding multiple heat shock proteins (average 164±220 TPM), the

antioxidant glutathione peroxidase (average 66±93 TPM), envelope stress-associated RNA polymerase sigma factor RpoE (average 37±67 TPM; Ades, 2004), a cold shock domain-containing protein (average 34±TPM), heat stress-associated RpoH (Nonaka *et al.*, 2006), and the stringent response-associated RNA polymerase binding protein DksA (Lemke *et al.*, 2011) were among the thirty five most abundant in symbiont-related transcriptomes (Figure 3.12).

Besides thioautotrophy and mixotrophy, other functions common to *C. orbiculata* symbiont species and other previously characterized thioautotrophic lucinid symbiont species (König *et al.*, 2016; Petersen *et al.*, 2016; Chapter II) included hydrogenotrophy (average 1±2 TPM), ammonia uptake (average 0.4±0.7 PM), denitrification (average 0.5±1 TPM), assimilatory nitrate reduction (average 0.03±0.04 TPM) and diazotrophy (average 0.08±0.1 TPM; Figure 3.13-3.15). *Candidatus* Thiodiazotropha endoloripes (Petersen *et al.*, 2016) and *Ca.* Sedimenticola endophacoides (Chapter II) could potentially hydrolyze urea, but only two of eight unbinned *C. orbiculata* gill metagenomes contained two genes homologous to urease subunit gamma related to that from Firmicutes species and urea ABC transporter substrate-binding protein related to that from *Methylomonas* species. Ten transcript clusters encoding allophanate hydrolase, urea carboxylase, urease accessory proteins and subunits of the urea ABC transporter were detected in two gill metatranscriptomes of OTU1-dominated specimens at average 0.02±0.03 TPM and two metatranscriptomes of OTU2-dominated specimens at average 0.1±0.1 TPM.

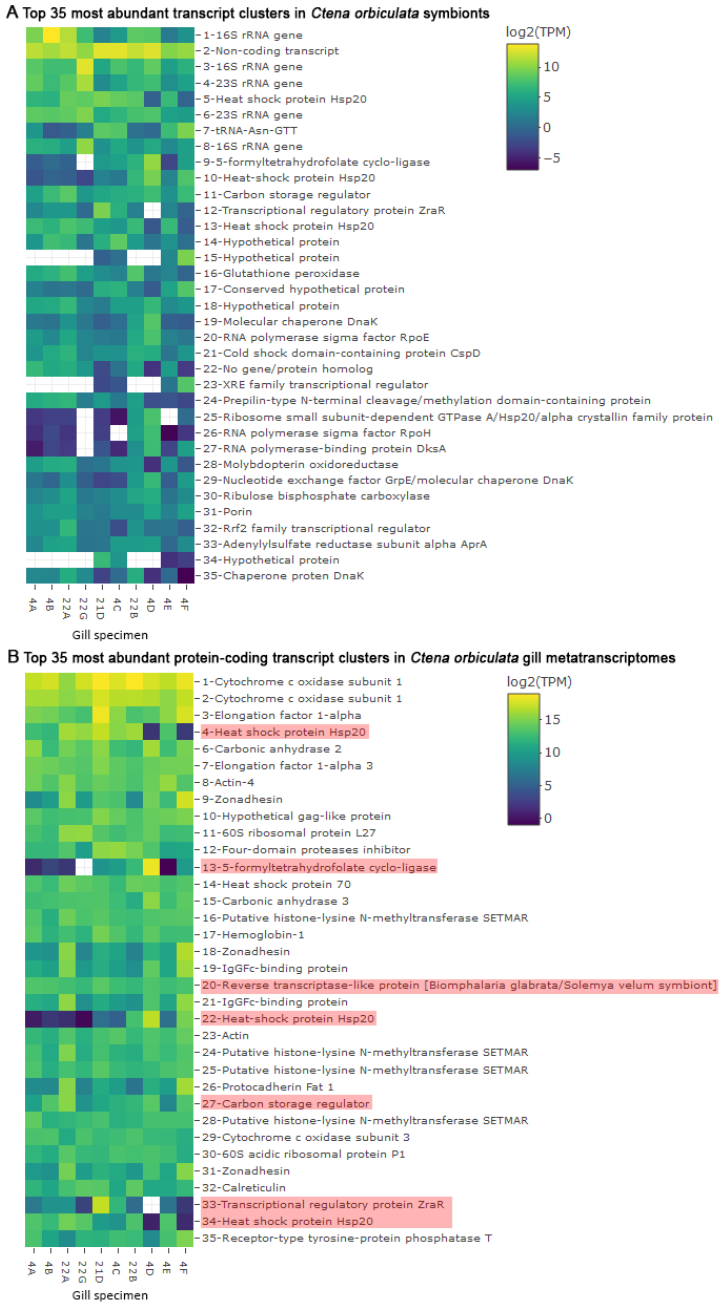


Figure 3.12. Log_2-transformed TMM-normalized TPM of gene products of the 35 most abundantly expressed (A) transcript clusters mapped to *C. orbiculata* symbionts and (B) protein-coding transcript clusters in sequenced gill metatranscriptomes. Bacteria-related transcripts in (B) are highlighted in pink.

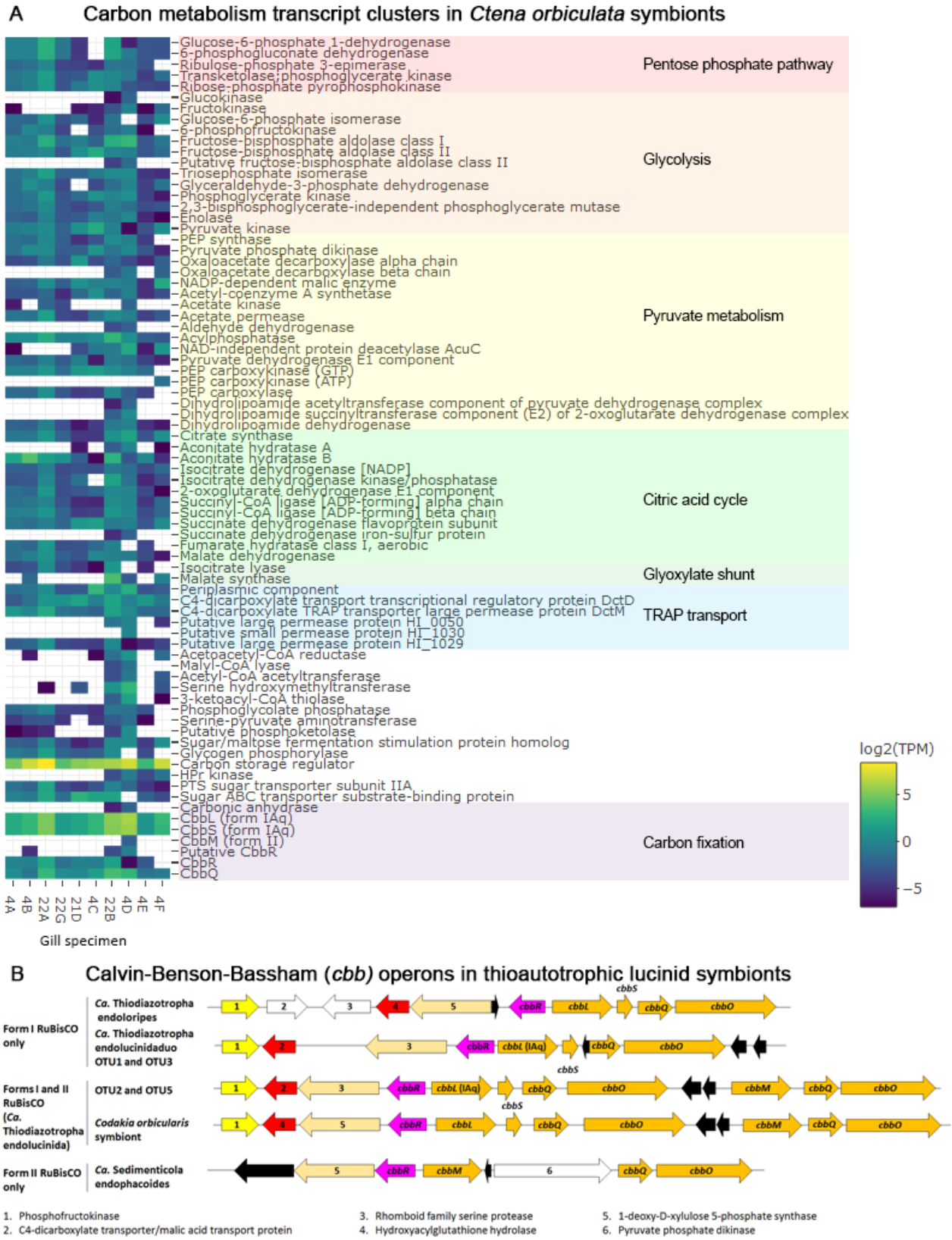

A
Carbon metabolism transcript clusters in Ctena orbiculata symbionts
Glucose-6-phosphate 1-dehydrogenase
6-phosphogluconate dehydrogenase
Ribulose-phosphate 3-epimerase
Transketolase;phosphoglycerate kinase
Ribose-phosphate pyrophosphokinase
Pentose phosphate pathway
Glucokinase
Fructokinase
Glucose-6-phosphate isomerase
6-phosphofructokinase
Fructose-bisphosphate aldolase class I
Fructose-bisphosphate aldolase class II
Putative fructose-bisphosphate aldolase class II
Triosephosphate isomerase
Glyceraldehyde-3-phosphate dehydrogenase
Phosphoglycerate kinase
2,3-bisphosphoglycerate-independent phosphoglycerate mutase
Enolase
Pyruvate kinase
Glycolysis
PEP synthase
Pyruvate phosphate dikinase
Oxaloacetate decarboxylase alpha chain
Oxaloacetate decarboxylase beta chain
NADP-dependent malic enzyme
Acetyl-coenzyme A synthetase
Acetate kinase
Acetate permease
Aldehyde dehydrogenase
Acylphosphatase
NAD-independent protein deacetylase AcuC
Pyruvate dehydrogenase E1 component
PEP carboxykinase (GTP)
PEP carboxykinase (ATP)
PEP carboxylase
Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex
Dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex
Dihydrolipoamide dehydrogenase
Pyruvate metabolism
Citrate synthase
Aconitate hydratase A
Aconitate hydratase B
Isocitrate dehydrogenase [NADP]
Isocitrate dehydrogenase kinase/phosphatase
2-oxoglutarate dehydrogenase E1 component
Succinyl-CoA ligase [ADP-forming] alpha chain
Succinyl-CoA ligase [ADP-forming] beta chain
Succinate dehydrogenase flavoprotein subunit
Succinate dehydrogenase iron-sulfur protein
Fumarate hydratase class I, aerobic
Malate dehydrogenase
Citric acid cycle
Isocitrate lyase
Malate synthase
Glyoxylate shunt
Periplasmic component
C4-dicarboxylate transport transcriptional regulatory protein DctD
C4-dicarboxylate TRAP transporter large permease protein DctM
Putative large permease protein HI_0050
Putative small permease protein HI_1030
Putative large permease protein HI_1029
TRAP transport
Acetoacetyl-CoA reductase
Malyl-CoA lyase
Acetyl-CoA acetyltransferase
Serine hydroxymethyltransferase
3-ketoacyl-CoA thiolase
Phosphoglycolate phosphatase
Serine-pyruvate aminotransferase
Putative phosphoketolase
Sugar/maltose fermentation stimulation protein homolog
Glycogen phosphorylase
Carbon storage regulator
HPr kinase
PTS sugar transporter subunit IIA
Sugar ABC transporter substrate-binding protein
Carbonic anhydrase
CbbL (form IAq)
CbbS (form IAq)
CbbM (form II)
Putative CbbR
CbbR
CbbQ
Carbon fixation
log2(TPM)
5
0
−5
4A
4B
22A
22G
21D
4C
22B
4D
4E
4F
Gill specimen
B
Calvin-Benson-Bassham (cbb) operons in thioautotrophic lucinid symbionts
Form I RuBisCO only
Ca. Thiodiazotropha endoloripes
Ca. Thiodiazotropha endolucinidaduo OTU1 and OTU3
Forms I and II RuBisCO (Ca. Thiodiazotropha endolucinida)
OTU2 and OTU5
Codakia orbicularis symbiont
Form II RuBisCO only
Ca. Sedimenticola endophacoides
cbbS
cbbR
cbbL
cbbQ
cbbO
cbbL (IAq)
cbbM
1. Phosphofructokinase
2. C4-dicarboxylate transporter/malic acid transport protein
3. Rhomboid family serine protease
4. Hydroxyacylglutathione hydrolase
5. 1-deoxy-D-xylulose 5-phosphate synthase
6. Pyruvate phosphate dikinase

Figure 3.13. (A) Log_2-transformed TMM-normalized TPM of gene products of autotrophy and heterotrophy-related transcript clusters mapped to *C. orbiculata* symbionts. (B) Comparison of the Calvin-Benson-Bassham (*cbb*) operon structures in *C. orbiculata* and other thioautotrophic lucinid symbionts (König *et al.*, 2016; Petersen *et al.*, 2016). White cells in (A) represent transcript clusters with zero TPM values. Black arrows in (B) depict genes encoding hypothetical proteins, colored arrows depict genes conserved in at least two species, and white arrows depict non-conserved genes. Abbreviations: PEP, phosphoenolpyruvate; NADP, nicotinamide adenine dinucleotide phosphate; NAD, nicotinamide adenine dinucleotide; AcuC, acetoin utilization protein; GTP, guanosine triphosphate; ATP, adenosine triphosphate; ADP, adenosine diphosphate; Dct, dicarboxylate transport proteins; TRAP, tripartite ATP-independent periplasmic transport; Hpr, histidine-containing phosphocarrier protein; PTS, phosphotransferase system.

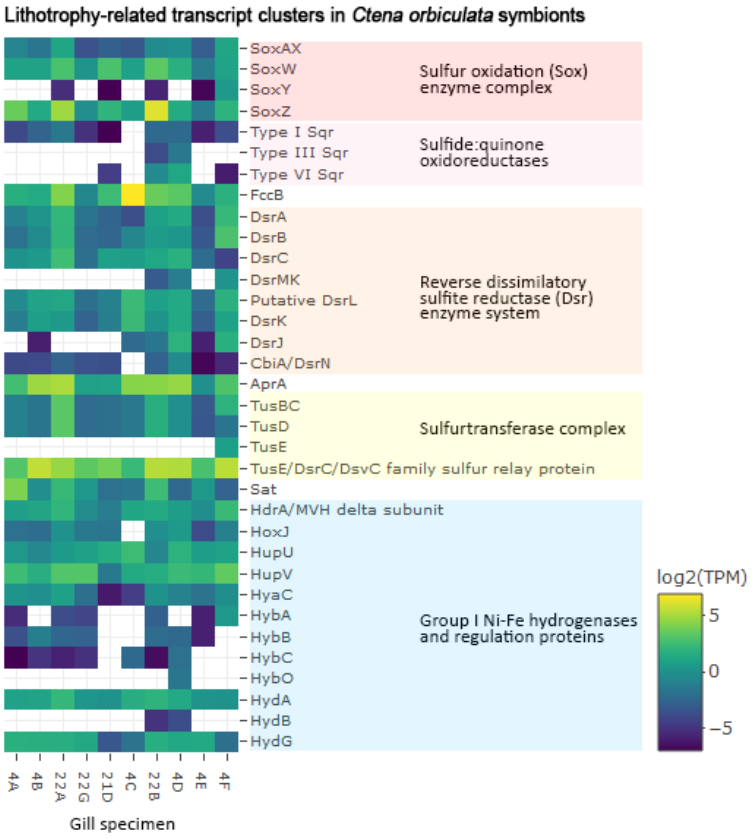


Figure 3.14. Log_2-transformed TMM-normalized TPM of gene products of lithotrophy-related transcript clusters mapped to *C. orbiculata* symbionts. White cells represent transcript clusters with zero TPM values. Abbreviations: FccB, sulfide dehydrogenase [flavocytochrome c] flavoprotein chain; CbiA, cobyrinate a,c-diamide synthase; AprA, adenylylsulfate reductase subunit alpha; Tus, tRNA 2-thiouridine synthesizing protein; Sat, sulfate adenylyltransferase; Tus, sulfur carrier proteins homologous to some Dsr proteins; HdrA/MVH, heterodisulfide reductase/methylviologen reducing hydrogenase; HoxJ, hydrogen uptake histidine-kinase; Hup, regulatory uptake hydrogenase; HyaC, membrane-bound Ni-Fe-hydrogenase I cytochrome b subunit; Hyb membrane-bound Ni-Fe hydrogenase 2; Hyd, periplasmic Ni-Fe hydrogenase.

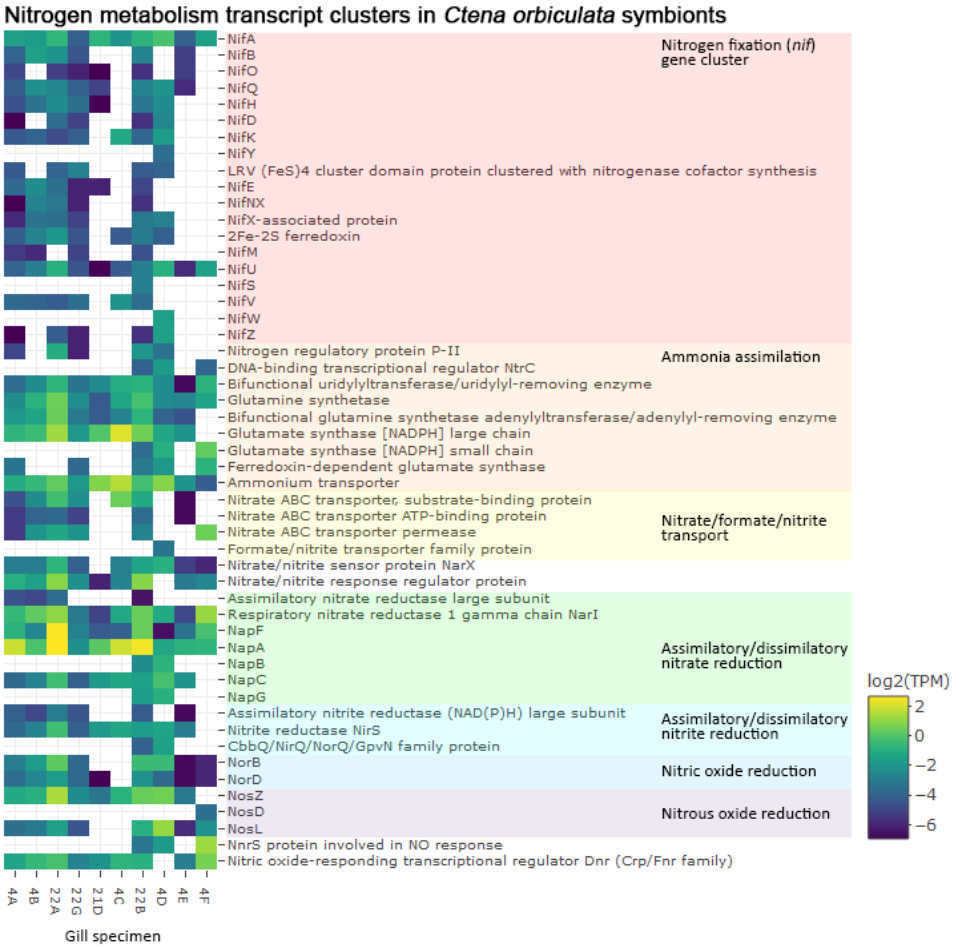


Figure 3.15. Log2-transformed TMM-normalized TPM of gene products of nitrogen metabolism-related transcript clusters mapped to *C. orbiculata* symbionts. White cells represent transcript clusters with zero TPM values. Abbreviations: Nif, nitrogen fixation proteins; LRV, leucine-repeat variant; NADPH, nicotinamide adenine dinucleotide phosphate; ABC, ATP-binding cassette; Cbb, proteins encoded by the Calvin-Bassham-Benson cycle operon; GpvN; gas vesicle protein; Nor, nitric oxide reductase; Nos, nitrous oxide reductase; Nnr, nitrite and nitric oxide reductase; Crp, cyclic adenosine monophosphate receptor protein; Fnr, fumarate nitrate reductase regulator.

Ctena orbiculata symbionts expressed flagella-related genes (average 0.9±5 TPM), pilus-related genes (average 1±7 TPM), as well as genes associated with phosphate uptake (average 0.3±0.8 TPM), polyphosphate utilization (average 0.2±0.6 TPM) and iron uptake (average 0.2±0.4 TPM). Biosynthetic genes for all twenty essential amino acids and vitamins B1, B2, B6, B7 and B9 (Figure 3.16a) and type I, II and VI secretion system genes (Figure 3.16b) were also identified in the symbionts' MAGs and transcriptomes. *Ctena orbiculata* symbionts encoded and expressed genes for VgrG, but not Hcp and VasL, exclusive to the type VI secretion system 2 gene cluster (Speare *et al.*, 2018).

In addition to thiotrophy and hydrogenotrophy-related genes, a C1 oxidation gene cluster encoding proteins involved in pyrroloquinoline quinone (PQQ) synthesis (average 0.2±0.6 TPM), PQQ-dependent methanol oxidation (Mdh; average 3±3 TPM), tetrahydromethanopterin (H_4MPT)-dependent formaldehyde oxidation (average 0.2±0.6 TPM) were conserved in all sequenced *C. orbiculata* symbionts and *Ca.* Thiodiazotropha endolucinida (König *et al.*, 2016; Figure 3.17a-c). Downstream of this gene cluster, another formate oxidation gene cluster encoding NADH-quinone oxidoreductase subunit F and formate dehydrogenase alpha subunit (FdhA) were predicted in all *C. orbiculata* symbionts, *Ca.* Thiodiazotropha endolucinida (König *et al.*, 2016), and *Ca.* Thiodiazotropha endoloripes (Petersen *et al.*, 2016).

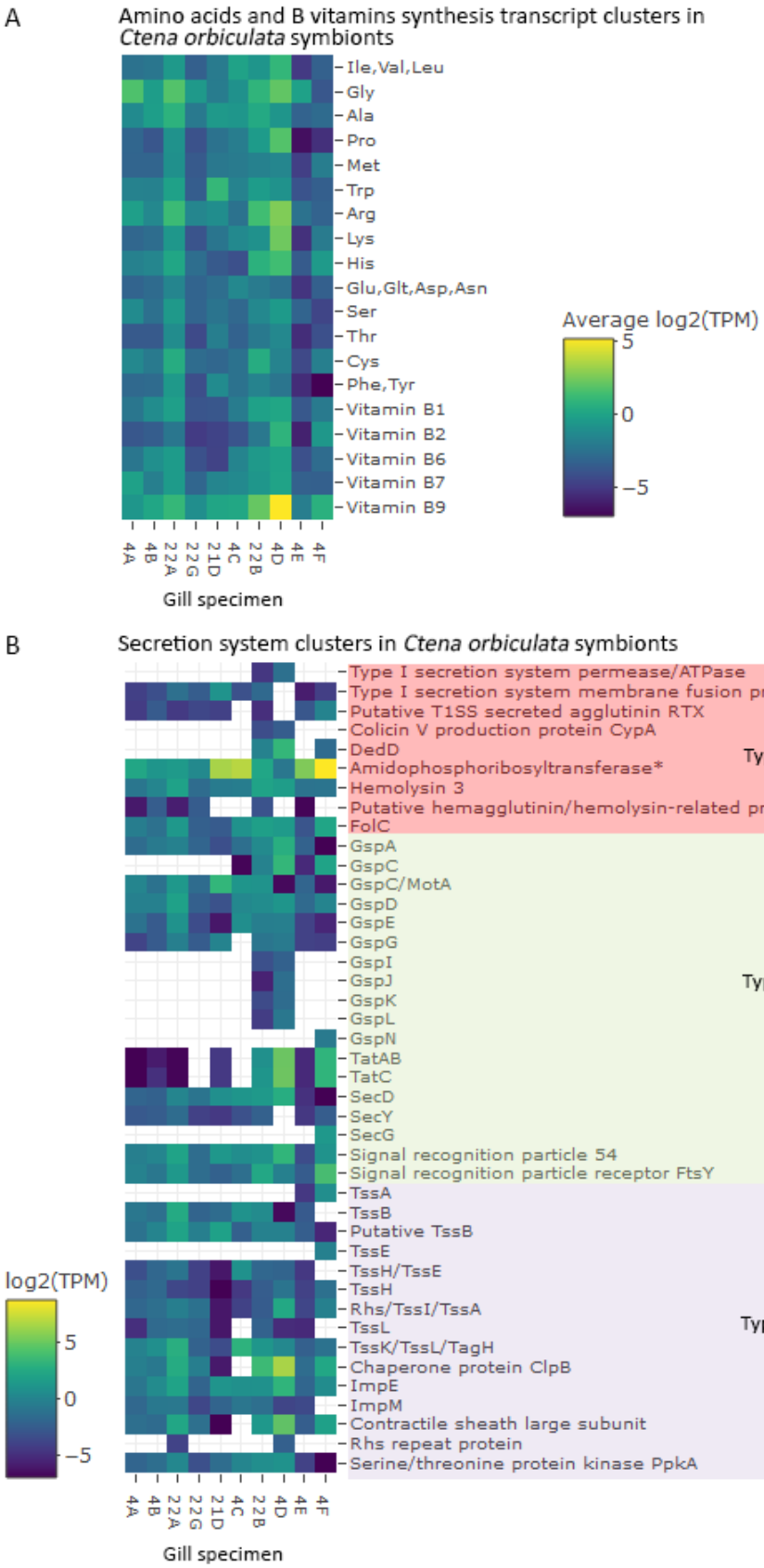

A
Amino acids and B vitamins synthesis transcript clusters in
Ctena orbiculata symbionts
Ile,Val,Leu
Gly
Ala
Pro
Met
Trp
Arg
Lys
His
Glu,Glt,Asp,Asn
Ser
Thr
Cys
Phe,Tyr
Vitamin B1
Vitamin B2
Vitamin B6
Vitamin B7
Vitamin B9
Average log2(TPM)
5
0
−5
4A
4B
22A
22G
21D
4C
22B
4D
4E
4F
Gill specimen
B
Secretion system clusters in Ctena orbiculata symbionts
Type I secretion system permease/ATPase
Type I secretion system membrane fusion protein PrsE
Putative T1SS secreted agglutinin RTX
Colicin V production protein CypA
DedD
Amidophosphoribosyltransferase*
Hemolysin 3
Putative hemagglutinin/hemolysin-related protein
FolC
Type I
GspA
GspC
GspC/MotA
GspD
GspE
GspG
GspI
GspJ
GspK
GspL
GspN
Type II
TatAB
TatC
SecD
SecY
SecG
Signal recognition particle 54
Signal recognition particle receptor FtsY
TssA
TssB
Putative TssB
TssE
TssH/TssE
TssH
Rhs/TssI/TssA
TssL
TssK/TssL/TagH
Chaperone protein ClpB
ImpE
ImpM
Contractile sheath large subunit
Rhs repeat protein
Serine/threonine protein kinase PpkA
Type VI
log2(TPM)
5
0
−5
4A
4B
22A
22G
21D
4C
22B
4D
4E
4F
Gill specimen

Figure 3.16. Log2-transformed TMM-normalized TPM of gene products of (A) amino acids (three-letter codes) and B vitamins biosynthesis and (B) secretion system transcript clusters mapped to *C. orbiculata* symbionts. White cells represent transcript clusters with zero TPM values. '*' indicates a multi-mapping transcript cluster. Abbreviations: RTX, repeats in toxin; DedD, cell division protein; FolC, dihydrofolate synthase/folylpolyglutamate synthase; Gsp, general secretory pathway protein; Tat, twin-arginine translocation proteins; Sec, secretory export proteins; Tss, type VI secretion system proteins; Tag, type VI secretion-associated protein; ClpB, caseinolytic peptidase B protein homolog; Imp; inner membrane protein.

Phylogenetic analyses of Mdh and FdhA protein sequences showed OTU- and species-specific clustering patterns across *C. orbiculata* symbionts (Figures 3.18-3.19) consistent with the phylogenomic tree (Figure 3.5). OTU4-related Mdh and FdhA sequences were the most closely related to those from *Ca.* Thiodiazotropha endolucinida (König *et al.*, 2016; Figures 3.18-3.19). Mdh sequences from lucinid symbionts clustered with sequences from the thioautotrophic gill symbiont of the giant Teredinidae bivalve *Kuphus polythalamia* (Distel *et al.*, 2017) and a marine purple sulfur bacterium *Thiorhodococcus drewsii* (Zaar *et al.*, 2003). These sequences formed a sister group with sequences from alphaproteobacterial species from the family Rhodospirillaceae, most of which were nitrogen-fixing (Figure 3.18). FdhA sequences from lucinid symbionts were most closely related to the free-living chemolithoautotrophic marine gammaproteobacterial species *Thioalbus denitrificans* (Park *et al.*, 2011), the *Kuphus polythalamia* symbiont

(Distel *et al.*, 2017), *Sedimenticola* spp. (Carlstrom *et al.*, 2015; Flood *et al.*, 2015) and methanotrophic gammaproteobacterial species (Figure 3.19). These sequences were in turn related to other nitrogen-fixing betaproteobacterial species (Figure 3.19). Mdh and FdhA protein sequences involved in methanol and formate oxidation showed distinct phylogeny. This suggests that the genes have a common origin in marine habitats and were acquired separately, possibly from diazotrophic alphaproteobacterial (Mdh) and betaproteobacterial species (FdhA). Two sets of qPCR primers targeting Mdh from OTU1 and OTU2 amplified matched DNA and cDNA from gill specimens dominated by these species. qPCR cDNA copy numbers of Mdh were consistent with TPM values observed in five out of seven amplified gill specimens (Figure 3.17).

Besides this gene cluster, other potentially C1 oxidation-related genes were also annotated in MAGs and/or transcriptomes of *C. orbiculata* symbionts. Genes homologous to deltaproteobacterial methanol:corrinoid methyltransferase, methanol methyltransferase corrinoid activation protein and methyltransferase corrinoid protein were identified in OTU2-related MAGs and transcriptomes (average 1±0.7 TPM), one OTU3-related the metatranscriptome (21D; 0.008 TPM; Figure 3.17c), and one OTU4-related unbinned assembly from specimen 4F. These genes potentially convert methanol to a corrinoid protein, which can be subsequently reduced to methane or oxidized to carbon dioxide (Abaibou *et al.*, 1995). OTU3-related MAGs and unbinned assemblies of all other *C. orbiculata* symbionts included the S-formylglutathione hydrolase (*estD*) gene homologous to free-living *Sedimenticola* species that converts S-formylglutathione to glutathione and formate in the glutathione-dependent pathway of formaldehyde detoxification (Chen *et al.*,

2016). *estD* was not expressed in the symbiont transcriptomes, but Mollusca-related *estD* was expressed in all gill metatranscriptomes at average 0.9±0.6 TPM. Additionally, OTU3-related MAGs contained oxalyl-CoA decarboxylase (*oxc*) and formyl-CoA transferase (*frc*) genes homologous to protein sequences in Betaproteobacteria and Alphaproteobacteria species. *oxc* converts oxalyl-CoA to formyl-CoA and CO_2, and *frc* converts formyl-CoA to formate and oxalyl-CoA (Azcarate-Peril *et al.*, 2006). *oxc* was not expressed in the metatranscriptomes, while a *frc* transcript cluster related to *Escherichia coli* was expressed in the OTU1-dominated gill specimen 22A (0.08 TPM) and the OTU2-dominated gill specimen 4D (0.3 TPM). Common C1-related genes identified in *C. orbiculata* and other lucinid symbionts encoded the bifunctional methylene-H_4F dehydrogenase/methenyltetrahydrofolate cyclohydrolase (FolD; average 0.2±0.2 TPM in *C. orbiculata* symbionts; not sequenced in *P. pectinatus* gill symbiont) and the respiratory formate dehydrogenase-O (average 1±2 TPM; Figure 3.17c; Abaibou *et al.*, 1995). *folD* likely participates in biosynthesis in these species because the tetrahydrofolate (H_4F)-dependent formaldehyde oxidation pathway was not completely sequenced in these symbionts (missing formyl-H_4F synthetase; Figure 3.17b; Vorholt, 2002).

C1 assimilation genes in the ribulose monophosphate (RuMP) pathway and many key genes in the serine-glyoxylate cycle (Smejkalova *et al.*, 2010) were not identified in the symbiont MAGs. Most of the fifteen accessory genes assigned by RAST (Aziz *et al.*, 2008) to the serine-glyoxylate cycle subsystem were also predicted in other carbon-related pathways (Table 3.5).

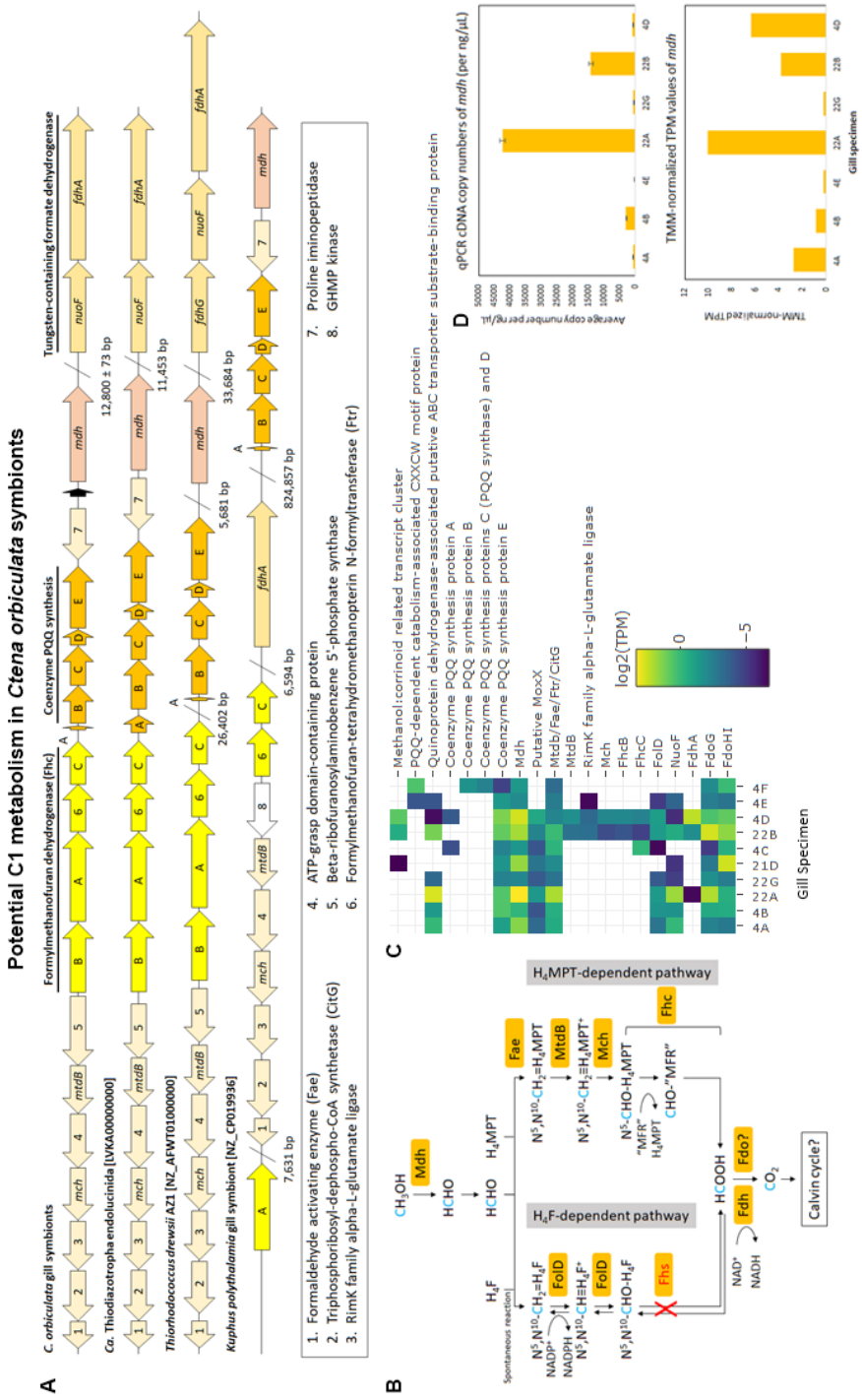


Figure 3.17. (A) Conserved gene clusters, (B) proposed pathways modified from (Vorholt, 2002; Pomper *et al.*, 2002), (C) TMM-normalized log_2TPM, and (D) qPCR copy numbers

and TMM-normalized TPM values of methanol dehydrogenase and/or other C1-oxidation genes in *C. orbiculata* symbionts. Colored arrows in (A) depict genes conserved in at least two species, and the black and white arrows represent genes encoding hypothetical and non-conserved proteins, respectively. Abbreviations: Mch, methenyl-tetrahydromethanopterin cyclohydrase; MtdB, NAD(P)-dependent methylene-tetrahydromethanopterin dehydrogenase; Mdh, pyrroloquinoline-quinone (PQQ)-dependent methanol dehydrogenase; NuoF, NADH-quinone oxidoreductase subunit F; Fdh, NAD-dependent tungsten-containing formate dehydrogenase; RimK, ribosomal protein S6 modification enzyme; GHMP, galacto-, homoserine, mevalonate and phosphomevalonate; H_4F, tetrahydrofolate; FolD, bifunctional methylene-H_4F dehydrogenase/methenyltetrahydrofolate cyclohydrolase; H_4MPT, tetrahydromethanopterin; Fhc, formyltransferase/hydrolase complex; Fdo, formate dehydrogenase O; Fae, formaldehyde activating enzyme; "MFR", postulated methanofuran analogue; Fhs, formyl-H_4F synthetase; ABC, ATP-binding cassette; MoxX, methanol utilization control regulatory protein.

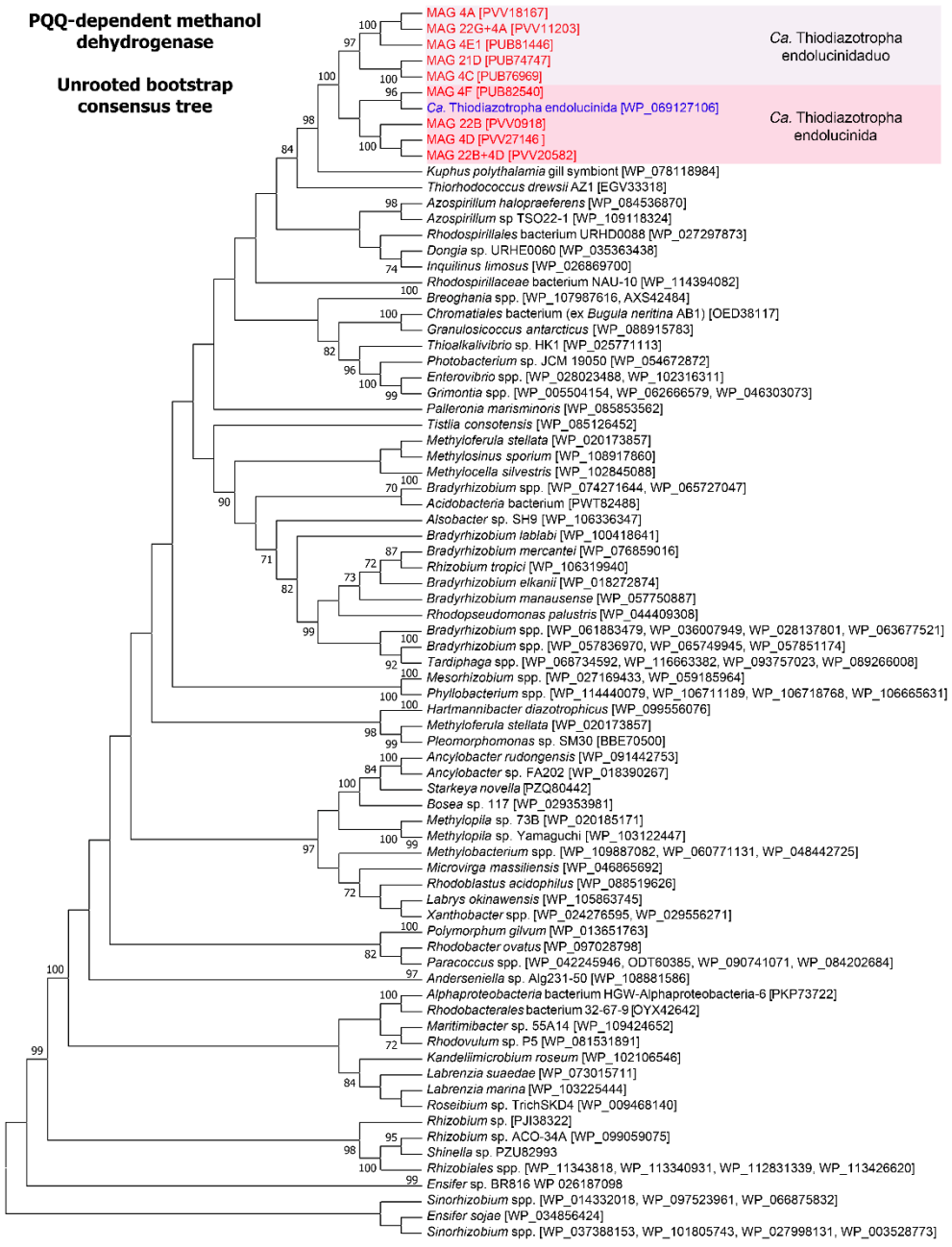


Figure 3.18. Unrooted bootstrap consensus maximum likelihood tree of methanol dehydrogenase protein sequences from *C. orbiculata* (red) in relation to other lucinid symbionts (blue) and other bacterial species. Tree nodes show bootstrap values and square brackets contain GenBank (Benson *et al.*, 2014) accession numbers.

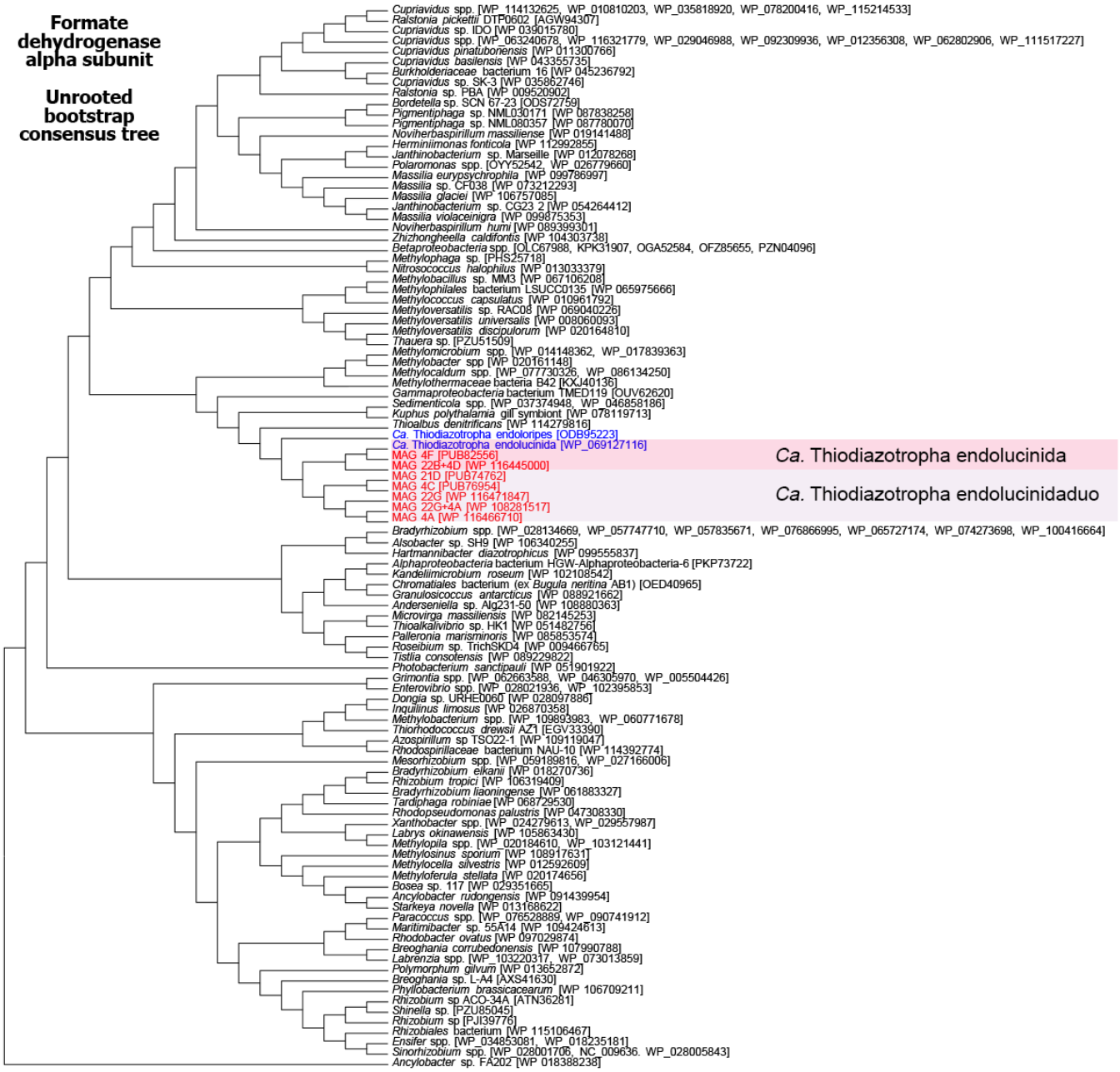


Figure 3.19. Unrooted bootstrap consensus maximum likelihood tree of formate dehydrogenase alpha protein sequences from *C. orbiculata* (red) in relation to other lucinid symbionts (blue) and other bacterial species. Tree nodes show bootstrap values and square brackets contain GenBank (Benson *et al.*, 2014) accession numbers.

Table 3.5. Serine-glyoxylate cycle-related gene products annotated in *Ctena orbiculata* symbionts and *Ca.* Thiodiazotropha endolucinida (Konig *et al.*, 2016; left) and other metabolic pathways associated with these gene products (right).

Pathway/gene product	Overlapping pathway(s)
Serine-glyoxylate cycle	
Serine hydroxymethyltransferase	Biosynthesis
Enolase	Glycolysis
Malyl-CoA lyase	Photorespiration
Malate dehydrogenase	TCA cycle; glyoxylate cycle
Citrate (si)-synthase	TCA cycle; glyoxylate cycle
Aconitate hydratase A	TCA cycle; glyoxylate cycle
Isocitrate lyase (glyoxylate cycle)	Glyoxylate Cycle
Succinate dehydrogenase flavoprotein subunit	TCA cycle
Succinate dehydrogenase iron-sulfur protein	TCA cycle
Succinyl-CoA ligase alpha chain	TCA cycle
Succinyl-CoA ligase beta chain	TCA cycle
Fumarate hydratase	TCA cycle
Pyruvate-alanine-serine interconversions	
Serine-pyruvate aminotransferase/L-alanine:glyoxylate aminotransferase	Photorespiration
Ethylmalonyl-CoA pathway	
Acetoacetyl-CoA reductase	Polyhydroxybutyrate metabolism; acetyl-CoA fermentation to butyrate
3-ketoacyl-CoA thiolase/acetyl-CoA acetyltransferase	Phenylalkanoic acid degradation; archaeal lipids

Core host functions

Carbonic anhydrase transcript clusters mapped to molluscan species (average 231±149 TPM) and the sea lamprey *Petromyzon marinus* (average 140±86 TPM), together with a lucinid-related hemoglobin 1 transcript cluster (average 124±60 TPM), were the sixth, fifteenth and seventeenth most abundant protein-coding transcripts in the gill metatranscriptomes respectively (Figure 3.12b). In contrast, lucinid-related hemoglobins 2 (average 0.2±0.7 TPM) and 3 (average 0.8±2 TPM) were expressed at lower levels. Two Mollusca-related transcript clusters encoding IgGFc-binding proteins (average 117±144 TPM) were the nineteenth and twenty-first most abundantly expressed in the gills (Figure 3.12b). Among lysozyme-associated transcripts in *C. orbiculata* (average 0.9±2 TPM), one molluscan-related transcript cluster encoding lysozyme 3 was the most highly expressed (average 4±2 TPM). Bivalve-related transcript clusters homologous to urease (average 1±3 TPM), urease accessory proteins (average 0.2±0.2 TPM) and urease transporters (average 0.2±0.5 TPM) were also identified in all gill metatranscriptomes.

Symbiont strain/species differences

Ctena orbiculata symbiont MAGs and transcriptomes showed very little inter-strain and inter-species variation of complete or near-complete metabolic pathways. MAGs of OTU2, OTU4 and *Ca.* Thiodiazotropha endolucinida (Konig *et al.*, 2016), which likely belong to the same species, encoded an additional form II RuBisCO (besides form Iaq) expressed only in the OTU2-dominated gill specimen 4D (0.205 TPM; Figure 3.13). Form

II RuBisCO genes were not identified in the unbinned metagenomes of OTU1- and OTU3-dominated gill specimens. For aerobic respiration, *C. orbiculata* symbionts and other clade A lucinid symbionts (Konig *et al.*, 2016; Petersen *et al.*, 2016) potentially utilize cbb3 (average 4±5 TPM in the former) and aa3 terminal oxidases (average 5±9 TPM). Additionally, OTU2-related MAGs and transcriptomes contained genes for cytochrome bd ubiquinol oxidase (average 0.5±0.4 TPM) also detected in the unbinned assembly and metatranscriptome of OTU4-dominated gill specimen 4F (0.02 TPM) and the metatranscriptome of OTU3-dominated gill specimen 4C (0.3 TPM).

Ctena orbiculata symbionts likely utilize distinct types of clustered regularly-interspaced short palindromic repeats (CRISPR)-associated genes. OTU1-related MAGs encoded type I-MYXAN (*Myxococcus xanthus*) CRISPR-associated protein Cas6/Cmx6 expressed in two of four OTU1-dominated transcriptomes (average 0.02±0.001TPM) and one OTU2-dominated gill specimen (4D; 0.03 TPM). Genes encoding type II CRISPR-associated endonucleases Cas2, Cas6, Cas9 were predicted only in OTU4-related MAGs (not expressed), while one OTU3-related MAG (21D) encoded type III-B CRISPR module-associated proteins Cmr1 (0.5 TPM) and Cmr2-6 (not expressed). Inter-taxa genetic differences in CRISPR-Cas system types among *C. orbiculata* symbionts suggest variations in mechanisms of prokaryotic defense against foreign DNA (Makarova *et al.*, 2011). The CRISPR-Cas system is involved in host colonization (Veesenmeyer *et al.*, 2014), innate immune avoidance (Sampson *et al.*, 2013), intracellular growth (Gunderson and Cianciotto, 2013), and virulence (Louwen *et al.*, 2014) in other host-microbe interactions, but its role in marine symbiosis has not been investigated.

Differential expression (DE) analyses across *C. orbiculata* symbiont communities

Differential expression (DE) analyses on symbiont-related genes using four DE algorithms showed <20 total DE transcript clusters ($p<0.05$, ≥2 fold-change) for OTU2-OTU3, OTU2-OTU4 and OTU3-OTU4 community comparisons (Figure 3.20). OTU1-OTU3 comparisons showed four total upregulated and 88 downregulated genes in OTU1-dominated communities, OTU1-OTU2 comparisons showed 751 total upregulated and 403 downregulated genes in OTU1-dominated communities, and OTU1-OTU4 comparisons showed ten upregulated and 72 downregulated genes in OTU1-dominated communities (Figure 3.20).

Transcript clusters upregulated in OTU1-dominated symbiont communities compared to communities containing significant abundances of other symbiont taxa were involved in bacterial secretion (type VI secretion protein Rhs/TssL/TssA and protein translocase subunit SecD), the transport of sugar, molybdate and an unknown substrate, and other functions (Figure 3.21). In relation to other taxa, OTU2-related communities preferentially expressed transcript clusters encoding an efflux transporter of toxic substances (Anes *et al.*, 2015), a ribosomal small subunit maturation protein GTPase A (Goto *et al.*, 2011), the signaling molecule diguanylate cyclase facilitating biofilm formation and pathogenesis (Schirmer, 2016), 5-formyltetrahydrofolate cyclo-ligase regulating purines, thymidylate and methionine biosynthesis and one-carbon metabolism (Meier *et al.*, 2007), heat shock protein, RpoH, and DksA (Figure 3.21). Commonly predicted Gene Ontology (GO; $p<0.05$) terms commonly enriched in OTU2-related communities compared to OTU1-dominated communities were associated with

nucleotide/nucleoside binding, proteolysis, calcium-transporting ATPase activity, aerobic respiration and drug response (Table 3.6).

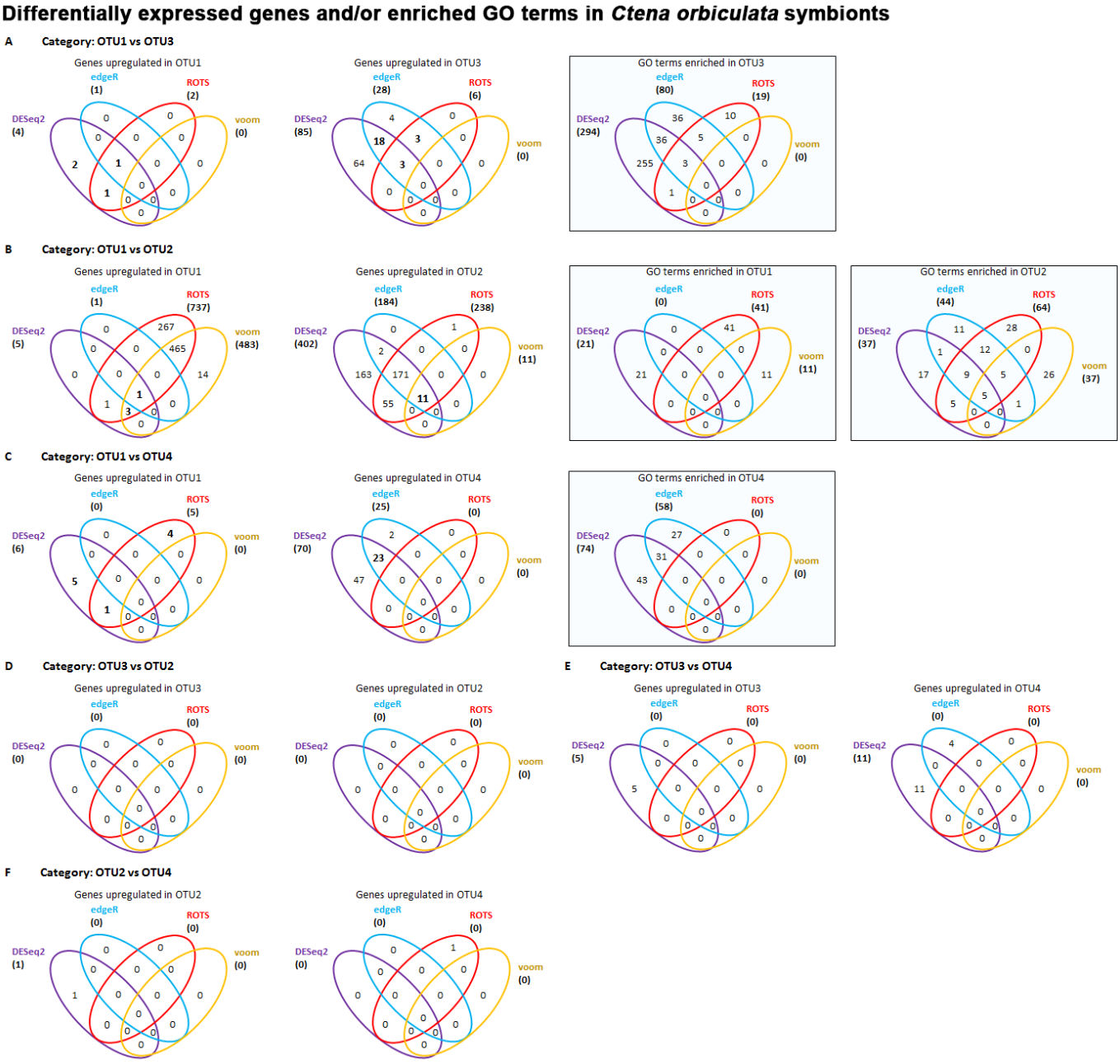


Figure 3.20. Venn diagrams of the numbers of differentially expressed (DE) genes ($p<0.05$, fold change≥ 2) predicted by four different algorithms across *C. orbiculata* symbiont taxa. Blue boxes show enriched gene ontology (GO) terms ($p<0.05$) within categories containing >20 total DE genes. Bold numerical text in Venn diagrams indicate the number of DE genes used to plot heatmaps in Figures 3.22-3.25.

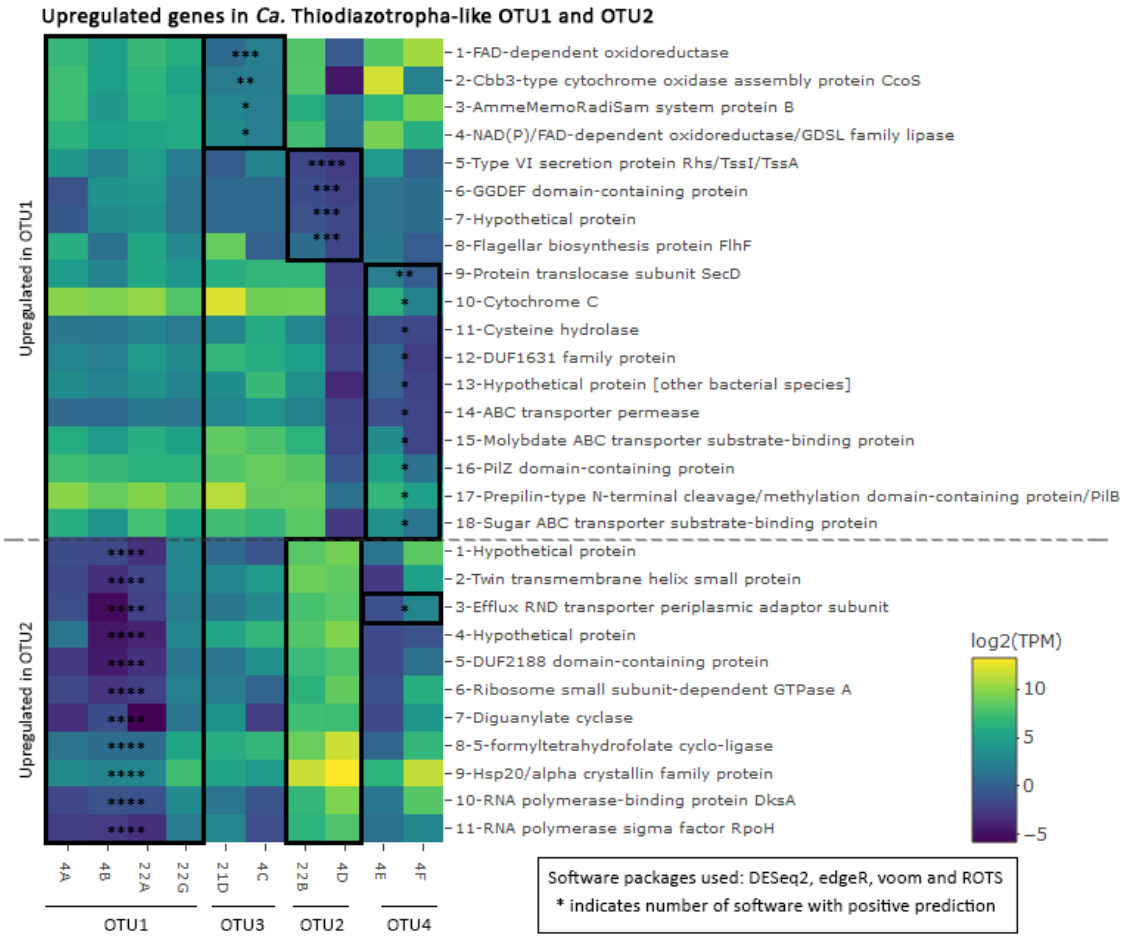


Figure 3.21. Upregulated genes ($p<0.05$, fold change≥ 2) predicted in OTU1 (above dashed line) and OTU2 (below dashed line) in relation to other *C. orbiculata* symbiont taxa (black boxes with asterisks). The number of asterisks indicate the number of DE software (four total) with positive predictions.

Table 3.6. List of deduplicated enriched Gene Ontology (Gene Ontology Consortium, 2015) terms ($p<0.05$) in transcriptomes of OTU2 compared to OTU1.

# software with positive predictions	Gene Ontology (GO) term	Mean over represented p-value	Mean under represented p-value
****	Guanyl ribonucleotide binding	0.01 ± 0.01	1 ± 0.001
****	Purine nucleoside/GTP binding	0.02 ± 0.01	1 ± 0.007
***	Endopeptidase activity	0.001 ± 0.0001	1 ± 0.004
**	Proteolysis	0.001 ± 0.02	1 ± 0.009
**	Peptidase activity	0.003	1 ± 0.0002
***	Calcium-transporting ATPase activity	0.004 ± 0.002	1 ± 0.001
***	Cyclic-di-GMP binding	0.005 ± 0.003	1 ± 0.004
***	Metalloendopeptidase activity	0.01 ± 0.002	1 ± 0.002
***	Calcium ion transmembrane transporter activity	0.01 ± 0.004	1 ± 0.003
***	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	0.01 ± 0.004	1 ± 0.003
***	Regulation of protein metabolic process	0.01 ± 0.002	1 ± 0.008
**	Peptidase activity, acting on L-amino acid peptides	0.01	1 ± 0.0001
**	Protein metabolic process	0.01	1 ± 0.001
**	Oxidative phosphorylation	0.01	1 ± 0.004
***	Protein tyrosine kinase activity	0.02 ± 0.01	1 ± 0.001
***	Aerobic electron transport chain	0.02 ± 0.01	1 ± 0.001
***	Ribonucleoside binding	0.02 ± 0.02	1 ± 0.007
**	Divalent inorganic cation transmembrane transporter activity	0.02	1 ± 0.001
**	Isoquinoline 1-oxidoreductase activity	0.02	1 ± 0.0004
**	Hydrogen:quinone oxidoreductase activity	0.02	1 ± 0
***	Single-stranded DNA 5'-3' exodeoxyribonuclease activity	0.03 ± 0.03	1 ± 0
**	Phenol-containing compound metabolic process	0.03	1 ± 0.001
**	Response to drug	0.03	1 ± 0.001
**	Peptide binding	0.03	1 ± 0.002
**	Metallopeptidase activity	0.03	1 ± 0.001
**	Cyclic nucleotide binding	0.04	1 ± 0.003
**	Purine ribonucleoside triphosphate metabolic process	0.04	1 ± 0.0003
**	Cellular respiration	0.04	1 ± 0.002
**	Protein catabolic process	0.05	1 ± 0.005
**	Aerobic respiration	0.05	1 ± 0.003
**	Intracellular membrane-bounded organelle	0.05	1 ± 0.005

Transcript clusters upregulated in OTU3-related communities compared to communities containing significant abundances of other taxa encoded a hypothetical conserved exported protein, flagella-related proteins, acetolactate synthase for branched chain amino acid synthesis (Chipman *et al.*, 1998), calmodulin, cytochrome c, a prevent-host-death protein promoting bacteriophage resistance and biofilm formation in *Pseudomonas* species (Petrova *et al.*, 2011), a nitrate ABC transporter and sulfide dehydrogenase (flavocytochrome c) flavoprotein subunit, among other proteins (Figure 3.22). GO terms commonly enriched in OTU3-related communities compared to OTU1-dominated communities were related to lysine biosynthesis, flagellar assembly, phosphatase activity, nucleotide binding, membrane components, protein/peptide secretion and nitrogen compound transport (Table 3.7). While flagellar genes have been implicated in symbiotic host attachment (Millikan and Ruby, 2004) and symbiont-to-host protein export (Maezawa *et al.*, 2006; Toft and Fares, 2008), their significance in the lucinid-bacteria symbiosis remains unknown. Similarly, the relevance of differentially expressed amino acid biosynthesis genes and their roles in shaping host-symbiont interdependencies on amino acids in this system remain to be elucidated.

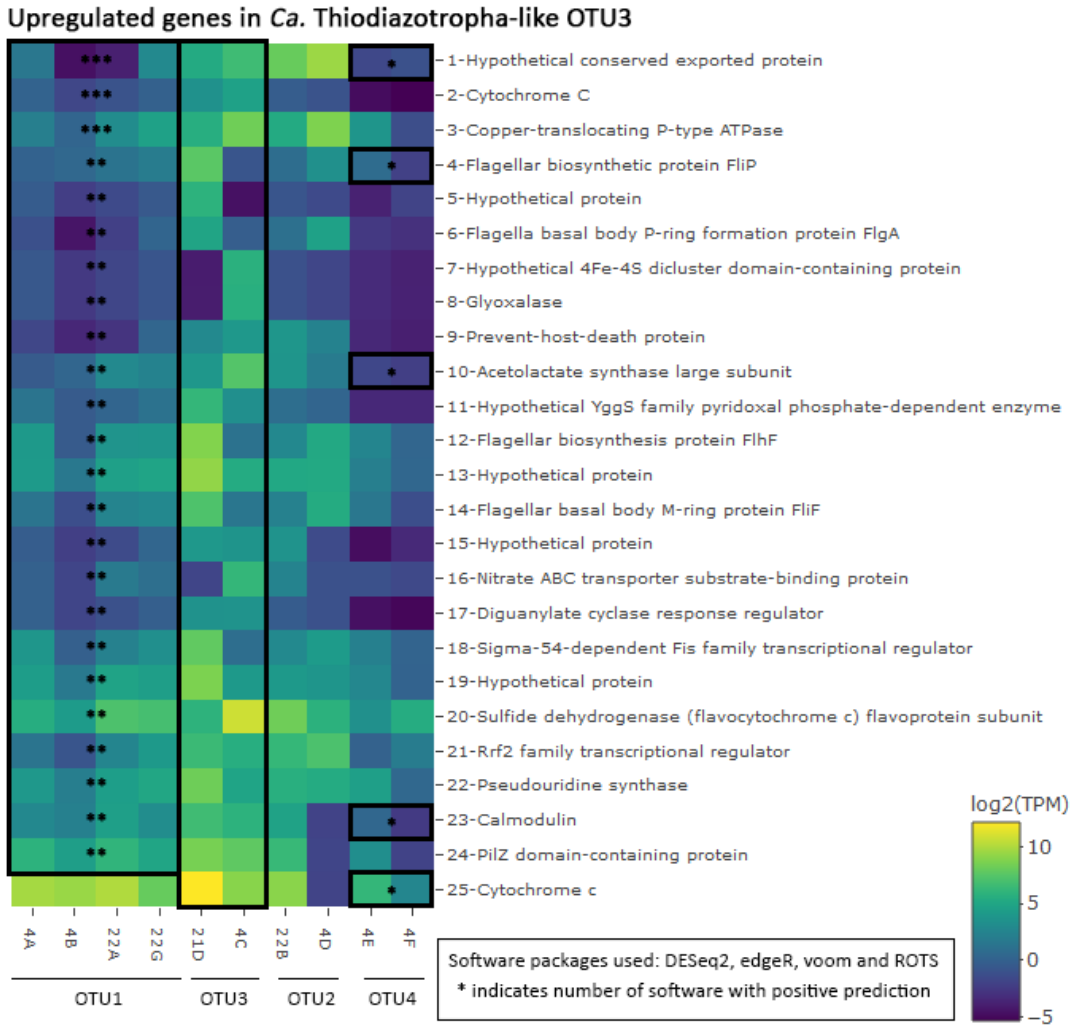


Figure 3.22. Upregulated genes ($p<0.05$, fold change≥ 2) predicted in OTU3 in relation to other *C. orbiculata* symbiont taxa (black boxes with asterisks). The number of asterisks indicate the number of DE software (four total) with positive predictions.

Table 3.7. List of deduplicated enriched Gene Ontology (Gene Ontology Consortium, 2015) terms ($p<0.05$) in transcriptomes of OTU3 compared to OTU1.

# software with positive predictions	Gene Ontology (GO) term	Mean over represented p-value	Mean under represented p-value
***	Transaminase activity	0.02 ± 0.009	1 ± 0.002
***	L,L-diaminopimelate aminotransferase activity	0.02 ± 0.02	1 ± 0
**	Bacterial-type flagellum organization	0.0005 ± 0.001	1 ± 0.00003
**	Cell projection organization	0.001 ± 0.002	1 ± 0.0001
**	Single-organism organelle organization	0.0007 ± 0.001	1 ± 0.00005
**	Organelle organization	0.005 ± 0.007	1 ± 0.001
**	Nucleoside-triphosphatase activity	0.02 ± 0.02	1 ± 0.006
**	Guanyl ribonucleotide binding	0.0005 ± 0	1 ± 0
**	Integral component of membrane	0.02 ± 0.03	1 ± 0.009
**	Intrinsic component of membrane	0.02 ± 0.03	1 ± 0.01
**	Pyrophosphatase activity	0.02 ± 0.03	1 ± 0.008
**	Hyrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	0.02 ± 0.03	1 ± 0.008
**	Protein/macromolecule localization	0.01 ± 0.01	1 ± 0.002
**	Secretion	0.02 ± 0.02	1 ± 0.002
**	Plasma membrane	0.01 ± 0.006	1 ± 0.002
**	Nitrogen compound transport	0.01 ± 0.02	1 ± 0.003
**	Membrane	0.02 ± 0.02	1 ± 0.008
**	Bacterial-type flagellum basal body, MS ring	0.01 ± 0.01	1 ± 0.00008
**	Protein/peptide/amide transport	0.01 ± 0.01	1 ± 0.001
**	Dicarboxylic acid metabolic process	0.004 ± 0.001	1 ± 0.0001
**	Establishment of protein localization	0.01 ± 0.01	1 ± 0.001
**	Peptide/protein secretion	0.02 ± 0.02	1 ± 0.001
**	Dicarboxylic acid biosynthetic process	0.02 ± 0.004	1 ± 0.001
**	Carboxy-lyase activity	0.03 ± 0.01	1 ± 0.002
**	Ornithine decarboxylase regulator/inhibitor activity	0.04 ± 0.02	1 ± 0.00004
**	Diaminopimelate decarboxylase activity	0.02 ± 0.01	1 ± 0.0001
**	Butanediol metabolic process; acetoin/secondary alcohol biosynthetic process	0.02 ± 0.009	1 ± 0
**	Molybdate ion transport/binding	0.02 ± 0.02	1 ± 0
**	Lysine metabolic/biosynthetic process via diaminopimelate	0.03 ± 0.03	1 ± 0.001

OTU4-related communities showed upregulation of transcript clusters encoding ribosomal proteins, RNA chaperone Hfq, BAX inhibitor of host apoptosis (Hemrajani *et al.*, 2010), cytochrome c oxidases involved in aerobic respiration, secretion (type VI secretion, twin-arginine translocation subunit TatA) and stress response (heat shock protein and DnaK; Figure 3.23). Accordingly, GO terms associated with these functions were enriched in OTU4-related communities compared to OTU1-dominated communities (Table 3.8). Upregulated host apoptosis-related symbiotic genes in OTU3-related communities and OTU4-related communities may be involved in the lucinid-bacteria symbiosis. For instance, *Wolbachia*-mediated inhibition of host apoptosis was proposed to effect host wasp transition from facultative parasitism to mutualism (Pannebakker *et al.*, 2007), but it is unclear whether similar mechanisms exist in the lucinid-bacteria symbiosis.

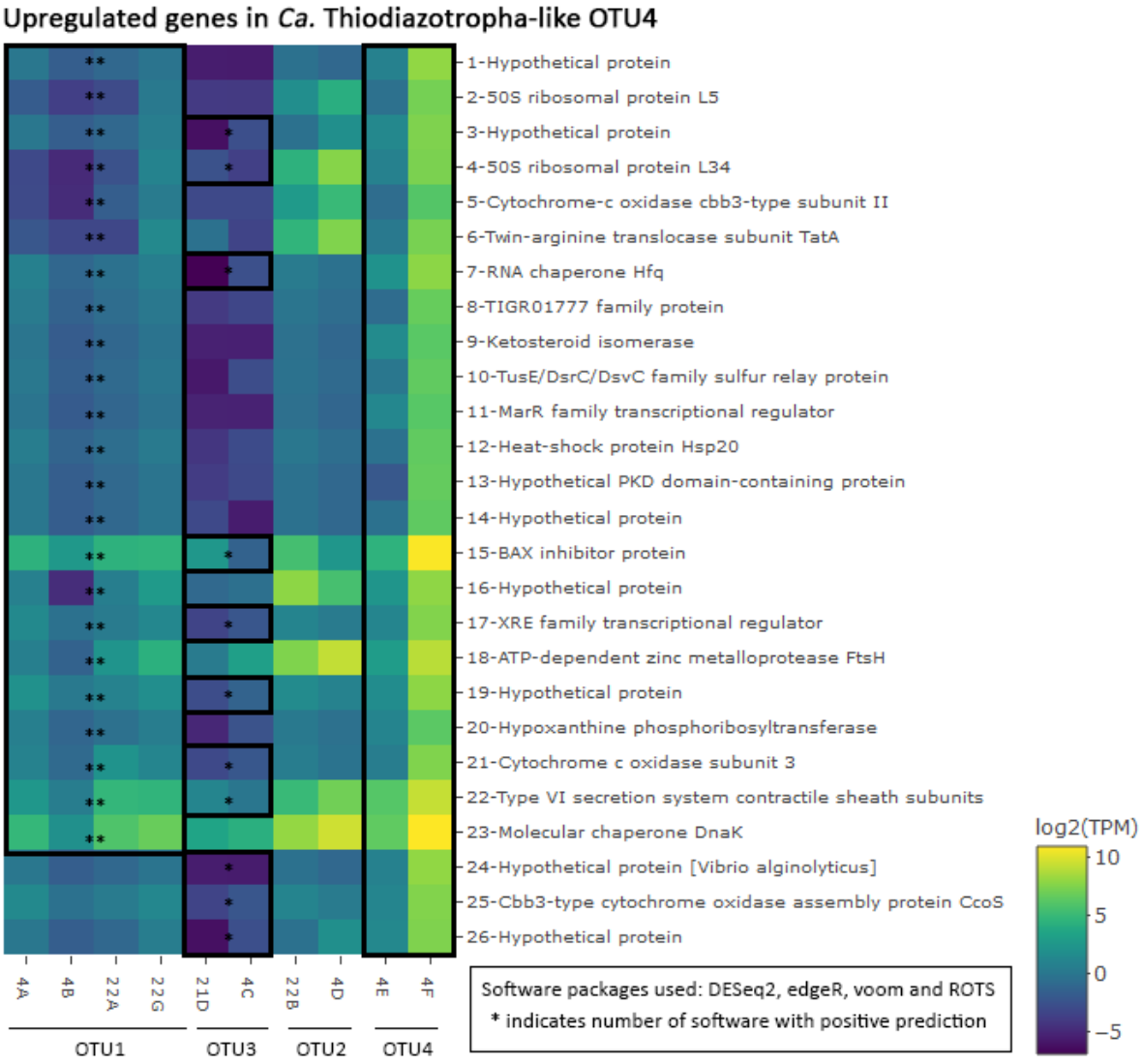


Figure 3.23. Upregulated genes ($p<0.05$, fold change≥ 2) predicted in OTU4 in relation to other *C. orbiculata* symbiont taxa (black boxes with asterisks). The number of asterisks indicate the number of DE software (four total) with positive predictions.

Table 3.8. List of deduplicated enriched Gene Ontology (Gene Ontology Consortium, 2015) terms ($p<0.05$) in transcriptomes of OTU4 compared to OTU1.

# software with positive predictions	Gene Ontology (GO) term	Mean over represented p-value	Mean under represented p-value
**	Aerobic electron transport chain	0.01 ± 0.007	1 ± 0.00001
**	Oxidoreductase activity, acting on a heme group of donors, oxygen as acceptor	0.01 ± 0.009	1 ± 0.00002
**	Oxidoreductase activity, acting on a heme group of donors	0.01 ± 0.01	1 ± 0.0006
**	Adenyl-nucleotide exchange factor activity	0.01 ± 0.009	1 ± 0
**	Nucleoside-triphosphatase/ATPase regulator activity	0.01 ± 0.009	1 ± 0
**	Ribosome	0.01 ± 0	1 ± 0.0006
**	Metalloendopeptidase activity	0.02 ± 0.02	1 ± 0.0015
**	Aerobic respiration	0.02 ± 0.02	1 ± 0.002
**	Chaperone binding	0.02 ± 0.01	1 ± 0.0002
**	Membrane insertase activity	0.02 ± 0.01	1 ± 0.0002
**	Structural constituent of ribosome	0.02 ± 0.003	1 ± 0.002
**	Aerobic respiration, using ferrous ions as electron donor	0.03 ± 0.02	1 ± 0.0002
**	Electron transporter, transferring electrons within cytochrome c oxidase complex activity	0.03 ± 0.02	1 ± 0.0002
**	Plasma membrane respiratory chain complex IV	0.03 ± 0.02	1 ± 0.0002
**	Serine-tRNA ligase activity	0.03 ± 0.02	1 ± 0.0002
**	Negative regulation of apoptotic process	0.03 ± 0.02	1 ± 0.0002
**	Selenocysteinyl-tRNA(Sec) biosynthetic process	0.03 ± 0.02	1 ± 0.0002
**	Phosphoribosyl-ATP diphosphatase activity	0.03 ± 0.02	1 ± 0.0003
**	Intracellular ribonucleoprotein complex	0.03 ± 0.005	1 ± 0.002
**	Protein transport by the Tat complex	0.03 ± 0.02	1 ± 0.0004
**	Respiratory electron transport chain	0.03 ± 0.006	1 ± 0.003
**	Translation	0.03 ± 0.006	1 ± 0.003
**	Cellular respiration	0.04 ± 0.008	1 ± 0.003
**	Endopeptidase activity	0.04 ± 0.01	1 ± 0.004

DE analyses of host-symbiont gene expression across quadrats

Besides taxon-specific clustering patterns, OTU1-dominated symbiont-specific transcriptomes collected from algae-covered and seagrass-covered quadrats (Figure 3.1) appeared to form two separate sub-clusters (Figure 3.7), although these transcriptomes shared 0.9±0.01 average pairwise PCC with each other. Exploratory DE analyses on these transcriptomes showed only five symbiont-related DE genes predicted by one (DESeq2; Love *et al.*, 2014) of four DE analysis algorithms used (Figure 3.24). Of these, the dissimilatory sulfite reductase transferase protein DsrC was upregulated in the algae-covered quadrat, while three transcript clusters encoding hypothetical proteins and one transcript cluster homologous to a cell wall-associated hydrolase from Alphaproteobacteria sp. were upregulated in the seagrass-covered quadrat (Figure 3.24). Host-related transcripts did not exhibit the same clustering patterns between algae-covered and seagrass-covered quadrats (Figure 3.9) but showed 73 putative DE genes predicted by both DESeq2 (Love *et al.*, 2014) and edgeR (Robinson *et al.*, 2010) across OTU-dominated metatranscriptomes in these quadrats (Figure 3.25). According to UniProt (The UniProt Consortium, 2015) annotations, these host-related DE genes were involved in a variety of muscle-related, cytoskeletal, co-chaperone, transcriptional, translational and other functions (Figure 3.25). Notably, host-related transcripts encoding cytochrome b-c1 complex, mitochondrial succinate dehydrogenase and mitochondrial sulfide:quinone oxidoreductase (Sqr) were predicted to be upregulated in the seagrass-covered quadrat (Figure S3.25).

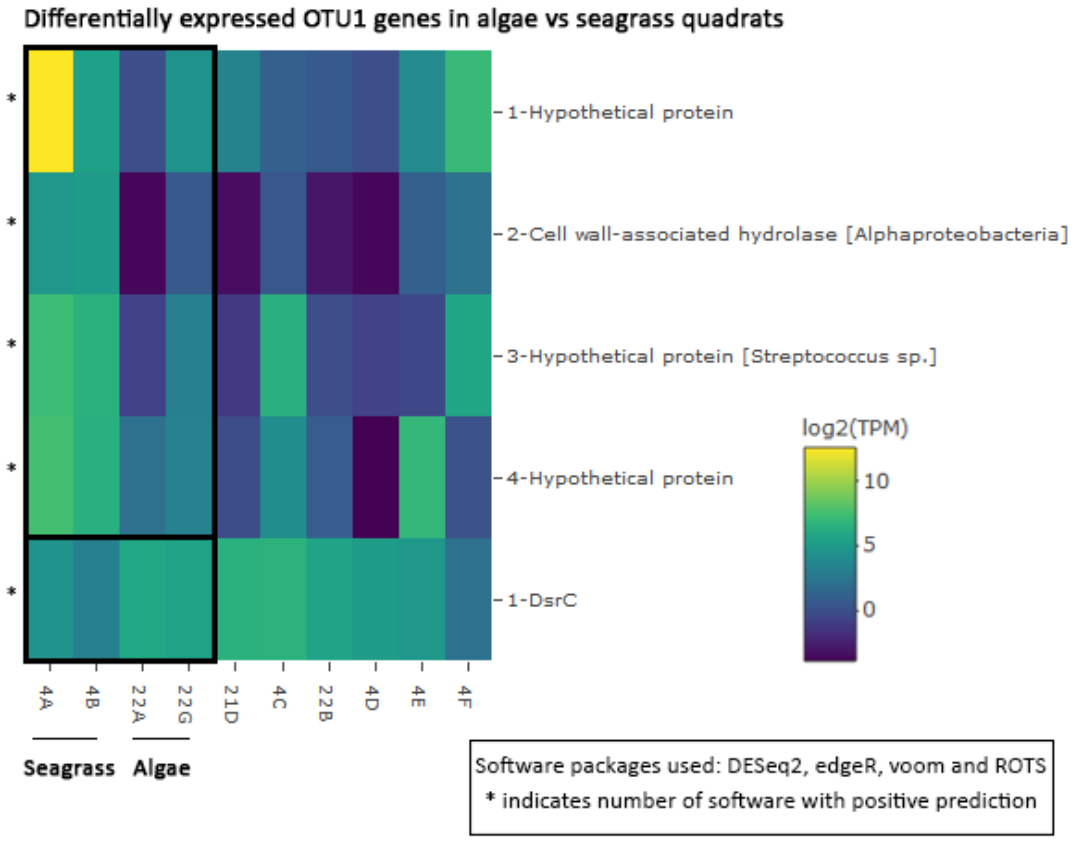
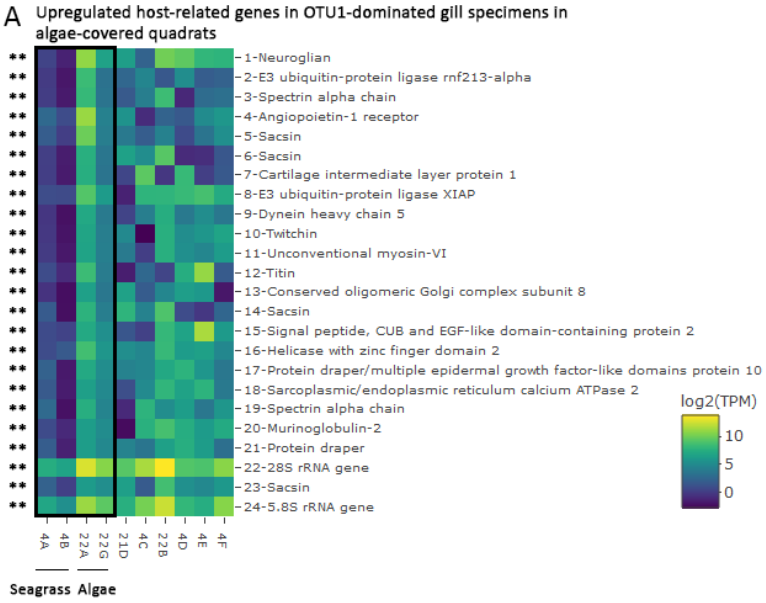
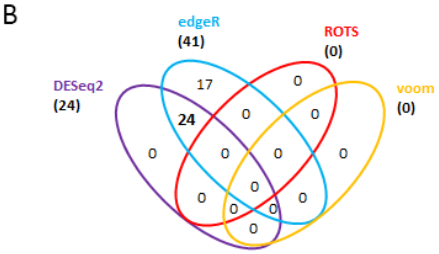
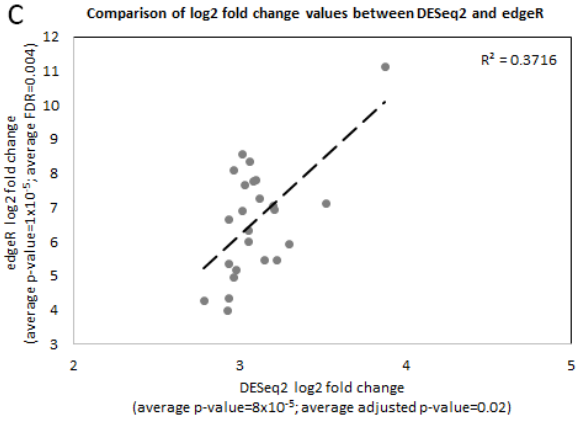
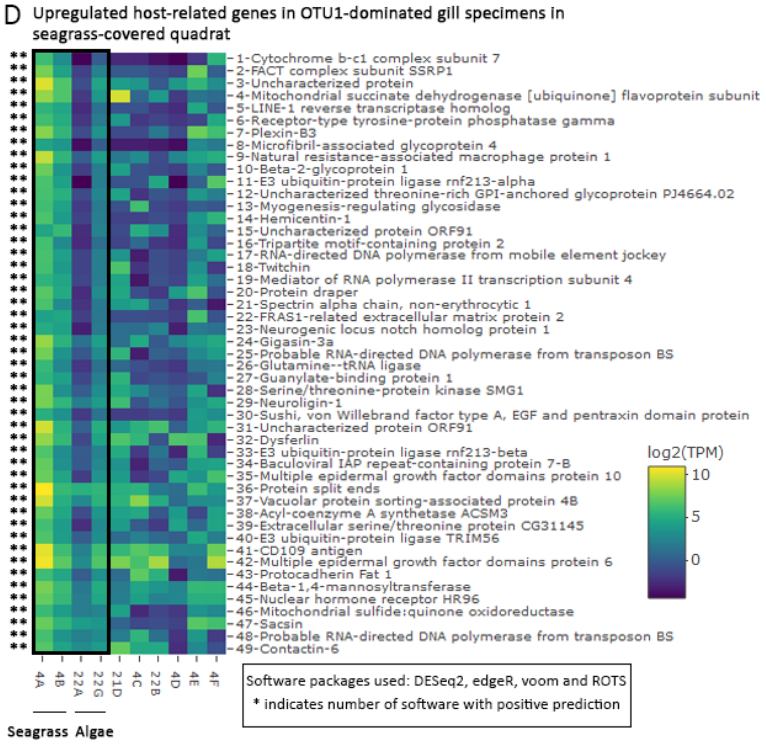


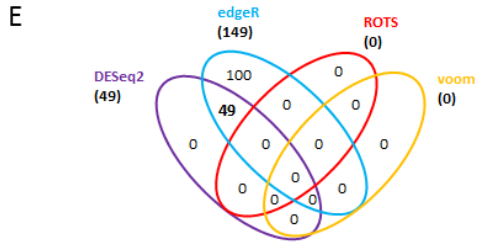
Figure 3.24. Differentially expressed ($p<0.05$, fold change≥ 2) genes mapped to OTU1-related MAGs between gill specimens collected from an algae-covered quadrat and those collected from a seagrass-covered quadrat (black boxes). All DE genes were predicted by the DESeq2 software package.


A
Upregulated host-related genes in OTU1-dominated gill specimens in algae-covered quadrats
1-Neuroglian
2-E3 ubiquitin-protein ligase rnf213-alpha
3-Spectrin alpha chain
4-Angiopoietin-1 receptor
5-Sacsin
6-Sacsin
7-Cartilage intermediate layer protein 1
8-E3 ubiquitin-protein ligase XIAP
9-Dynein heavy chain 5
10-Twitchin
11-Unconventional myosin-VI
12-Titin
13-Conserved oligomeric Golgi complex subunit 8
14-Sacsin
15-Signal peptide, CUB and EGF-like domain-containing protein 2
16-Helicase with zinc finger domain 2
17-Protein draper/multiple epidermal growth factor-like domains protein 10
18-Sarcoplasmic/endoplasmic reticulum calcium ATPase 2
19-Spectrin alpha chain
20-Murinoglobulin-2
21-Protein draper
22-28S rRNA gene
23-Sacsin
24-5.8S rRNA gene
log2(TPM)
10
5
0
4A 4B 22A 22G 21D 4C 22B 4D 4E 4F
Seagrass Algae


B
edgeR
(41)
ROTS
(0)
DESeq2
(24)
voom
(0)
17
24


C
Comparison of log2 fold change values between DESeq2 and edgeR
R² = 0.3716
edgeR log2 fold change
(average p-value=1x10⁻⁵; average FDR=0.004)
DESeq2 log2 fold change
(average p-value=8x10⁻⁵; average adjusted p-value=0.02)


D
Upregulated host-related genes in OTU1-dominated gill specimens in seagrass-covered quadrat
1-Cytochrome b-c1 complex subunit 7
2-FACT complex subunit SSRP1
3-Uncharacterized protein
4-Mitochondrial succinate dehydrogenase [ubiquinone] flavoprotein subunit
5-LINE-1 reverse transcriptase homolog
6-Receptor-type tyrosine-protein phosphatase gamma
7-Plexin-B3
8-Microfibril-associated glycoprotein 4
9-Natural resistance-associated macrophage protein 1
10-Beta-2-glycoprotein 1
11-E3 ubiquitin-protein ligase rnf213-alpha
12-Uncharacterized threonine-rich GPI-anchored glycoprotein PJ4664.02
13-Myogenesis-regulating glycosidase
14-Hemicentin-1
15-Uncharacterized protein ORF91
16-Tripartite motif-containing protein 2
17-RNA-directed DNA polymerase from mobile element jockey
18-Twitchin
19-Mediator of RNA polymerase II transcription subunit 4
20-Protein draper
21-Spectrin alpha chain, non-erythrocytic 1
22-FRAS1-related extracellular matrix protein 2
23-Neurogenic locus notch homolog protein 1
24-Gigasin-3a
25-Probable RNA-directed DNA polymerase from transposon BS
26-Glutamine--tRNA ligase
27-Guanylate-binding protein 1
28-Serine/threonine-protein kinase SMG1
29-Neuroligin-1
30-Sushi, von Willebrand factor type A, EGF and pentraxin domain protein
31-Uncharacterized protein ORF91
32-Dysferlin
33-E3 ubiquitin-protein ligase rnf213-beta
34-Baculoviral IAP repeat-containing protein 7-B
35-Multiple epidermal growth factor domains protein 10
36-Protein split ends
37-Vacuolar protein sorting-associated protein 4B
38-Acyl-coenzyme A synthetase ACSM3
39-Extracellular serine/threonine protein CG31145
40-E3 ubiquitin-protein ligase TRIM56
41-CD109 antigen
42-Multiple epidermal growth factor domains protein 6
43-Protocadherin Fat 1
44-Beta-1,4-mannosyltransferase
45-Nuclear hormone receptor HR96
46-Mitochondrial sulfide:quinone oxidoreductase
47-Sacsin
48-Probable RNA-directed DNA polymerase from transposon BS
49-Contactin-6
log2(TPM)
10
5
0
4A 4B 22A 22G 21D 4C 22B 4D 4E 4F
Seagrass Algae
Software packages used: DESeq2, edgeR, voom and ROTS
* indicates number of software with positive prediction


E
edgeR
(149)
ROTS
(0)
DESeq2
(49)
voom
(0)
100
49

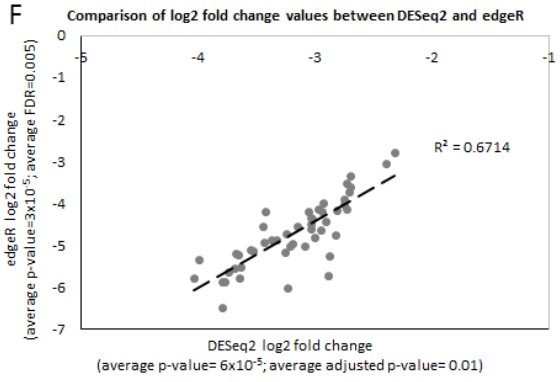

F
Comparison of log2 fold change values between DESeq2 and edgeR
R² = 0.6714
edgeR log2 fold change
(average p-value=3x10⁻⁵; average FDR=0.005)
DESeq2 log2 fold change
(average p-value= 6x10⁻⁵; average adjusted p-value= 0.01)

Figure 3.25. Differentially expressed (DE; $p<0.05$, fold change≥ 2) genes mapped to host-related genes between OTU1-dominated gill specimens (black boxes) collected from an algae-covered quadrat and those collected from a seagrass-covered quadrat (black boxes). (A) Upregulated host-related genes in algae-covered quadrat, (B) Venn diagram of the numbers of host-related DE genes upregulated in algae-covered quadrat predicted by four different algorithms, (C) correlation of $\log_2$ fold change values of host-related DE genes upregulated in algae-covered quadrat commonly predicted by DESeq2 (Love *et al.*, 2014) and edgeR (Robinson *et al.*, 2010), (D) Upregulated host-related genes in seagrass-covered quadrat, (E) Venn diagram of the numbers of host-related DE genes upregulated in seagrass-covered quadrat predicted by four different algorithms, (F) correlation of $\log_2$ fold change values of host-related DE genes upregulated in seagrass-covered quadrat commonly predicted by DESeq2 (Love *et al.*, 2014) and edgeR (Robinson *et al.*, 2010). Bold numerical text in Venn diagrams indicate the number of DE genes used to plot heatmaps in (A) and (D) and scatterplots in (C) and (F).

Discussion

To date, lucinid symbiont species and their associated functional variations remain largely unexplored because, out of >100 identified lucinid species listed on NCBI (NCBI Resource Coordinators, 2016), only three gill symbiont species from three host species have been comprehensively sequenced (Konig *et al.*, 2016; Petersen *et al.*, 2016; Chapter II). Despite previous marker-gene based diversity studies (Durand and Gros, 1996; Durand *et al.*, 1996; Brissac *et al.*, 2011; Brissac *et al.*, 2016) and a recent gill microbiome, metagenome and metatranscriptome characterization study on *P. pectinatus* (Chapter II), in-depth investigations into taxonomic, genetic and functional symbiont variations within a single lucinid host population are currently lacking in the literature. Within our study site (Sugarloaf Key, Florida, USA), we hypothesized that strain-level symbiont diversity might be present within a single host (*C. orbiculata*) population, which can be detected using high-resolution -omics methods. We discovered the co-existence of multiple thioautotrophic gill endosymbiont species and/or strains (OTUs 1-4) with marked genetic differences. These symbionts shared a large number of core genes and functions, which included a well-conserved, previously undiscovered C1 oxidation pathway. On the other hand, gill metatranscriptomic analyses revealed host-related functions that may contribute to symbiont metabolic support and symbiont aggregation, as well as host and/or symbiont genes differentially expressed across symbiont taxa and quadrats with different vegetation coverages.

Thioautotrophic symbiont taxonomic diversity observed in our studied *C. orbiculata* population contrast and parallel Brissac *et al.*'s findings of strain-level symbiont diversity, defined by haplotypes, within *C. orbiculata* and five other lucinid species in the Caribbean (Brissac *et al.*, 2016). While Brissac *et al.*'s lucinid samples showed a high degree of host-symbiont specificity and all their *C. orbiculata* samples were colonized by only a single symbiotic haplotype (Brissac *et al.*, 2016), we discovered much greater symbiont taxonomic diversity within the same host species at our collection site. Our results point to a lower degree of host-symbiont specificity than reported in Brissac *et al.*, likely because of either higher taxonomic symbiont diversity encountered by the host in the environment (Brissac *et al.*, 2011) or less stringent host regulation on symbiont acquisition in this studied *C. orbiculata* population (Brissac *et al.*, 2016). The observed symbiont diversity in *C. orbiculata* also supports the notion of a symbiont community (50), rather than a homogenous symbiont strain, associated with this *C. orbiculata* population. Similar marine symbiont communities have been reported, for example, in the light organs of the squids *Sepiola affinis and Sepiola robusta*, where closely related *Vibrio fischeri* and *V. logei* were detected at different abundances (Mushegian and Ebert, 2015). Species-level symbiont heterogeneity was also reported in *Osedax* gutless marine worms, which hosted two closely related Oceanospirillales heterotrophic symbiont species (Goffredi *et al.*, 2014). It is unlikely that the identified symbiont taxa are exclusive to *C. orbiculata*, because our preliminary analyses showed the *C. orbiculata* symbionts to share identical 16S rRNA gene sequences with symbionts in *Codakia orbicularis* from the same sampling site (unpublished data). Previous studies have also shown the lesser Antillean *C. orbiculata* to

share the same symbiont species with five other Caribbean lucinid species (Durand and Gros, 1996; Durand *et al.*, 1996; Gros *et al.*, 2003).

Few inter-taxa differences were observed across *Ctena orbiculata* symbiont taxa. *Candidatus* Thiodiazotropha endolucinidaduo (OTU1 and OTU3) and *Ca.* Thiodiazotropha endoloripes MAGs (Petersen *et al.*, 2016) encoded and expressed form Iaq RuBisCO, while *Ca.* Thiodiazotropha endolucinida spp. (OTU2, OTU5, and Konig *et al.*'s *Codakia orbicularis* symbionts; 2016) encoded and expressed form Iaq and form II RuBisCO. In contrast, only form II RuBisCO was predicted in MAG of the mangrove-associated *Phacoides pectinatus* symbiont, *Ca.* Sedimenticola endophacoides (Chapter II). Compared to form II RuBisCO, form Iaq RuBisCO is more efficient at distinguishing between the competing substrates oxygen and CO_2 (Tabita *et al.*, 2008). Inter-clade variations in the RuBisCO variants used by lucinid symbionts suggest the possibility of symbiont clade-specific divergent evolution and/or gene duplication, which may be due to varying intracellular and/or extracellular oxygen levels experienced by the common ancestor of each symbiont clade either during their host-associated or free-living stage. For aerobic respiration, seagrass-associated lucinid symbionts (Konig *et al.*, 2016; Petersen *et al.*, 2016), including *C. orbiculata* symbionts, can potentially use the high-affinity cbb3 type and low-affinity aa3 type terminal oxidases under low and high oxygen concentrations (García-Horsman *et al.*, 1994; Pitcher and Watmough, 2004), respectively. Additional, a cytochrome bd ubiquinol oxidase adapted to microaerobic environments (Borisov *et al.*, 2011) was also encoded and expressed mainly by *Ca.* Thiodiazotropha-like OTU2. This enzyme may allow for adaptive responses to variable oxygen intracellular and/or

extracellular environments or facilitate symbiotic nitrogen fixation via oxygen scavenging and respiratory protection, as demonstrated in other free-living and plant-associated nitrogen-fixing bacteria (Kaminski *et al.*, 1996; Poole and Hill, 1997; Dincturk *et al.*, 2011). Glutathione peroxidase-related transcripts highly expressed in *C. orbiculata* symbionts may also confer similar protective functions by scavenging H_2O_2, as shown in the legume-Rhizobia symbiosis (Bianucci *et al.*, 2017). Differentially expressed genes across *C. orbiculata* symbiont taxa encoded secretion, stress-response, transport, and biosynthesis proteins possibly relevant to the lucinid-bacteria symbiosis. However, the statistical significance of DE analysis in this study is limited because of the lack of robust replicates consisting of pure symbiotic monocultures.

Despite their taxonomic diversity, thioautotrophic *C. orbiculata* symbionts shared a high number of common genes and functions. These included previously characterized lithoautotrophy, diazotrophy and potential heterotrophy functions (Konig *et al.*, 2016; Petersen *et al.*, 2016; Chapter II). Additionally, many transcripts involved in temperature, oxidative, envelope and nutrition stress responses were highly expressed in *C. orbiculata* symbionts and likely reflect stresses from intracellular host selection mechanisms (Chapter II) parallel to the *Eupyrmna-Vibrio* symbiosis (Weis *et al.*, 1996; Small and McFall-Ngai, 1999; Davidson *et al.*, 2004) or stresses caused by the external environment. Defense-related transcripts involved in type I, II and VI secretion systems, such as colicin-related transcripts, general secretion (Sec)-related transcripts twin-arginine translocation (Tat)-related transcripts may reduce symbiont-symbiont competition by killing closely related strains (Cascales *et al.*, 2007; Chapter II) and contribute to host infection (Green and

Mecsas, 2016; Nivaskumar and Francetic, 2014). Besides known lithotrophic pathways, genes for the oxidation of C1 compounds including methanol, formaldehyde and formate were conserved in MAGs of *C. orbiculata* symbionts and discovered in this study in *Ca.* Thiodiazotropha endolucinida. *Ctena orbiculata* symbionts may use C1 compounds as an energy source, because they encoded and expressed formate dehydrogenase O enabling the use of formate as an electron donor during respiration (Abaibou *et al.*, 1995). In the quadrats sampled in this study, dissolved methane was detected at low concentrations between 0.3 to 22 μg/L in the sampled quadrats. Concentrations of methanol, formaldehyde and formate were not measured. Methane and methanol could be present in the study site at low levels as by-products released by plants from unknown pathways (methane) or cell pectin demethylation in plant cell walls (methanol; Nemecek-Marshall *et al.*, 1995; Keppler *et al.*, 2006). While free-living methanotrophs demonstrated high methane consumption activity in aquatic plants, including algae and seagrasses (Yoshida *et al.*, 2014), methanol was observed to be phytotoxic and mutualistic interactions between methylotrophic bacteria and strawberry plants and seagrasses have been proposed (Abanda-Nkpwatt *et al.*, 2006; Crump *et al.*, 2018). Based on previous reports, we hypothesize that C1 oxidation in *C. orbiculata* symbionts potentially benefit their surrounding algae and seagrasses through methanol detoxification. Because there is no substantial genetic evidence for RuMP and serine-glyoxylate C1 assimilation pathways in these symbiont species, we speculate that the CO_2 end-product of C1 oxidation could either be fixed via the autotrophic Calvin-Benson-Bassham cycle, as demonstrated in the

diazotrophic alphaproteobacterial *Xanthobacter* strain 25a (Croes *et al.*, 1991), or used to conserve energy.

Analyses of host-related transcripts have also revealed both similarities and differences in the host metabolism of *C. orbiculata* compared to previously described *P. pectinatus* (Chapter II). While *P. pectinatus* harboring non-diazotrophic symbionts showed consistently high expression levels of hemoglobins 1, 2 and 3, *C. orbiculata* only showed high expression of the sulfide-reactive hemoglobin 1 (Kraus and Wittenberg, 1990). In contrast, expression levels of oxygen-reactive (Kraus and Wittenberg, 1990) hemoglobins 2 and 3 were at least 100x lower than that of hemoglobin 1 in *C. orbiculata*. Parallel to the host, we also observed high expression of antioxidant glutathione peroxidase in the *C. obiculata* symbionts. In contrast, porewater dissolved oxygen concentrations in the sampled quadrats were within the range of oxygen concentrations previously measured in a sub-oxic *Lucinoma aequizonata* habitat (Cary *et al.*, 1989; Hentschel *et al.*, 1993). These suggest that, in comparison to *P. pectinatus*, *C. orbiculata* and its symbionts experience higher levels of oxygen in their intracellular, but not extracellular, environments and/or require tighter host-symbiont co-operative regulation on intracellular oxygen concentrations to facilitate nitrogen fixation. *Phacoides pectinatus* and *C. orbiculata* both expressed lysozyme transcripts, presumably for bactericidal activity (Chapter II), although expression levels of lysozyme transcripts in the latter were lower in relation to other host-related transcripts. Strong expression of molluscan-related IgGFc-binding protein-coding transcripts was also observed in *C. orbiculata*, but not yet observed in other lucinid species. These transcripts are potentially involved in mucosal defense (Kobayashi *et al.*, 2002) and

may facilitate initial symbiont selection of tightly aggregated symbionts like in the *Eupyrmna-Vibrio* symbiosis (Nyholm and McFall-Ngai, 2003).

Unlike previous studies on symbionts of Caribbean lucinid species and the *Sepiola* squids, where host geographic location (Brissac *et al.*, 2016) and temperature (Nishiguchi, 2000) influenced the observed species distribution, respectively, we did not recognize any apparent spatial trend explaining the taxonomic distribution of *C. orbiculata* symbionts in the four sampled quadrats. We did, however, note that two gill specimens containing OTU4 most closely related to Brissac *et al.*'s *C. orbiculata* symbiont haplotype (Brissac *et al.*, 2016) occurred only in one single seagrass-dominated quadrat. Based on the observed symbiont taxonomic distribution, we speculate that symbiont acquisition by clams in this population could be based on random encounters, as suggested by previous research on lucinid species in the Phillippines (Brissac *et al.*, 2011) and on bacterial communities associated with the green macroalga *Ulva australis* (Burke *et al.*, 2011).

Our metatranscriptomic analyses also revealed potential clustering patterns of OTU1-related transcripts collected from algae-covered and seagrass-covered quadrats, although very few symbiotic genes, including thiotrophy-related DsrC (Cort *et al.*, 2008), were predicted to be preferentially expressed in the algae-covered quadrat. Host-related transcripts did not follow the same quadrat-specific clustering pattern, but showed candidate up-regulated genes encoding cytochrome b-c1 complex involved in aerobic respiration and oxidative stress-triggered apoptosis (Dibrova *et al.*, 2013), mitochondrial succinate dehydrogenase that connects the tricarboxylic cycle to the electron transport chain (Van Vranken *et al.*, 2015), and mitochondrial Sqr that oxidizes sulfide to thiosulfate

(Marcia *et al.*, 2010) in seagrass-covered compared to algae-dominated quadrats. These transcripts are of particular relevance to the tripartite symbiosis model between seagrass, lucinid clams and their symbionts, which is based on symbiotic sulfide detoxification, host-symbiont aerobic respiration (Reynolds *et al.*, 2014; van der Heide *et al.*, 2012), and possibly nitrogen fixation (Petersen *et al.*, 2016). Nevertheless, our differential expression analyses were hindered by limitations on sample sizes and the number of replicates.

Overall, this study uncovered taxonomic, genetic and functional thioautotrophic gill endosymbiont diversity in *C. orbiculata* and furthers our current understanding of host-symbiont specificity, physiology and interactions. Our findings highlight the intriguing, poorly understood complexity of lucinid-bacteria symbioses and generate a range of new testable hypotheses encompassing the establishment, persistence, stability, and distribution of symbiont communities; the significance of one-carbon metabolism in thioautotrophic lucinid symbionts; the roles of other bacterial taxa in lucinid symbioses; and three-way interactions between the environment, lucinid hosts and their symbionts. Future studies, such as cross-infection experiments, imaging experiments, controlled aquarium experiments, and large-scale field studies coupled with -omics analyses would continue to elucidate the range of host-symbiont functions across species, environmental gradients, and habitats possible in this remarkably diverse symbiotic system.

CHAPTER THREE

STRUCTURE AND FUNCTIONS OF GILL MICROBIOME SPECIES IN THE SYMBIOTIC COASTAL LUCINID BIVALVE *STEWARTIA FLORIDANA*

Authors

Shen Jean Lim[1], Brenton G. Davis[1,2], Danielle E. Gill[1], John Swetenburg[3], Annette Summers Engel[4], Laurie C. Anderson[5], and Barbara J. Campbell[1]

[1]Department of Biological Sciences, Clemson University, Clemson, SC 29634-0001

[2]Current address: College of Medicine, Medical University of South Carolina, Charleston, SC 29425-8900

[3]Current address: University of South Carolina School of Medicine Greenville, Greenville, SC 29605-4208

[4]Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN 37996-1410

[5]Department of Geology and Geological Engineering, South Dakota School of Mines and Technology, Rapid City, SD 57701-3901

Authors' contributions

A.S.E., B.J.C., and L.C.A. secured the funding for this study, supervised sample collection, and research efforts; S.J.L., B.J.C., A.S.E., L.C.A. collected the samples used in the study; S.J.L. and B.J.C. conceived the experiments; S.J.L. performed most of the experiments, software implementation, data analyses, and wrote this chapter. B.G.D.

performed qPCR on the thioautotrophic symbiont; D.E.G. performed qPCR on the *Spirochaeta*-like species; J.S. prepared metatranscriptomic libraries for *S. floridana* gill specimens; S.J.L. maintains the NCBI sequence data and L.C.A. curates the metadata and maintains specimens of dissected tissues and valves. B.J.C reviewed and edited this chapter.

Introduction

All extant bivalve species from the family Lucinidae host thioautotrophic gammaproteobacterial endosymbionts in specialized gill bacteriocytes (Taylor and Glover, 2000). These symbionts are related to a large clade of diverse chemosynthetic marine symbionts that mainly produce organic matter for lucinid hosts using energy derived from the oxidation of sulfur, hydrogen, and possibly C1 compounds (Dubilier *et al.*, 2008; Petersen *et al.*, 2016; König *et al.*, 2016; Chapters II and III). Among lucinid species, thioautotrophic gammaproteobacterial symbionts are clustered into three distinct clades within a paraphyletic group. Clade A is associated with bivalves in seagrass or sulfide-poor environments, whereas clade B is associated with mangrove-dwelling *Anodontia* spp., and clade C are found in mangrove-dwelling *Phacoides pectinatus* (Cavanaugh *et al.*, 2006; Brissac *et al.*, 2011; Chapter II). The lucinid-bacteria chemosymbiosis has an ancient history, likely originating from the Silurian period (Liljedahl, 1992), with diversification of coastal lucinid clams being attributed to the emergence of mangroves and seagrasses in the late Cretaceous period (Stanley, 2014). Today, besides a variety of shallow marine habitats, lucinids are also found in deep sea hydrothermal vents (Glover *et al.*, 2004), cold

seeps (Brissac *et al.*, 2011), and mud volcanos (Rodrigues *et al.*, 2010). Lucinids colonize oxic-anoxic interfaces of their habitats, where they acquire oxygen from oxic waters, as well as CO_2, H_2, reduced sulfur compounds, nitrogenous compounds, and, in some cases, C1 compounds, from anoxic sediment pore waters to support their physiology (Dando *et al.*, 1994; Taylor and Glover, 2000) and respiratory, lithotrophic, and/or diazotrophic functions of their gill symbionts (Dando *et al.*, 1994; Hentschel *et al.*, 1996; Taylor and Glover, 2000; Duplessis *et al.*, 2004; Stewart *et al.*, 2005; Kleiner *et al.*, 2015; König *et al.*, 2016; Petersen *et al.*, 2016; Chapters II and III). Lucinid symbionts in seagrass-dwelling hosts also improve seagrass health in an interdependent, tripartite symbiosis by detoxifying sulfide, fixing carbon, and, in some cases, fixing nitrogen and detoxifying methanol levels (Petersen *et al.*, 2016; Chapter III), in return for host and symbiont access to oxygen from seagrass roots (Fisher and Hand, 1984; Johnson *et al.*, 2002; van der Heide *et al.*, 2012; Reynolds *et al.*, 2014; Stanley, 2014).

To date, studies on lucinid bivalves have focused on their paleontology (Liljedahl, 1992; Taylor and Glover, 2009; Anderson, 2014; Taylor and Glover, 2018), ecology and taxonomy (Williams *et al.*, 2004; Glover *et al.*, 2004; Oliver and Holmes, 2006; Taylor *et al.*, 2011), physiology (Frenkiel *et al.*, 1997; Duplessis, Dufour *et al.*, 2004; van der Geest *et al.*, 2014; Christo *et al.*, 2016), gill and bacteriocyte morphology (Distel and Felbeck, 1987; Liberge *et al.*, 2001; Ball *et al.*, 2009), as well as mode of environmental symbiont acquisition (Gros *et al.*, 1996; Gros *et al.*, 1999; Gros *et al.*, 2003; Brissac *et al.*, 2009; Gros *et al.*, 2012; Espinosa *et al.*, 2013). Studies focused on lucinid symbionts describe their range of metabolic functions (Felbeck *et al.*, 1981; Hentschel *et al.*, 1996; König *et*

al., 2016; Petersen *et al.*, 2016; Chapters II and III), as well as phylogeny and diversity (Durand and Gros, 1996; Durand *et al.*, 1996; Duperron *et al.*, 2007; Brissac *et al.*, 2016; Chapter III). However, the complexity of the lucinid gill microbiomes is only beginning to be comprehensively examined (Chapters II and III), although the possible presence of other non-thioautotrophic gill-associated bacterial taxa in lucinid gills has been noted for over ten years (Ball *et al.*, 2009; Duperron *et al.*, 2012; Espinosa *et al.*, 2013). The first of these studies describe 16S rRNA gene phylotypes belonging to the thioautotrophic symbiont and another *Spirochaeta*-like species in the gills of *Lucinoma aff. kazani* (Duperron *et al.*, 2007). A subsequent imaging study on the gills of *Euanodontia ovum* show the presence of extracellular spirochete-like bacteria (8–10 μm length and 0.2–0.3 μm width), intracellular clade B thioautotrophic symbionts (3–5 μm length and 0.5–1.0 μm width), and another intracellular rod-shaped bacterium (1 μm length and 0.4–0.5 μm width; Ball *et al.*, 2009). Two 16S rRNA gene phylotypes related to the thioautotrophic symbiont and another gammaproteobacterial species outside of the chemosynthetic marine symbiont clade have also been identified in the gills of *Loripes orbiculatus* (Espinosa *et al.*, 2013). More recently, in-depth metagenomic and metatranscriptomic profiling of the *Phacoides pectinatus* gill microbiome reveals the presence and functions of three bacterial species, including the thioautotrophic symbiont species, a heterotrophic gammaproteobacterial *Kistimonas*-like species, and a heterotrophic *Spirochaeta*-like species (Chapter II). Similar -omics characterization of the gill microbiome in *Ctena orbiculata* also detect thioautotrophic symbiont species and strain-level heterogeneity, as well as non-thioautotrophic *Endozoicomonas*-like species (Chapter III).

Despite previous research efforts, hypotheses and conclusions on lucinid gill microbiome diversity and functions center on a small number of mostly seagrass-associated taxa. Furthermore, ecosystem-based studies integrating habitat geochemistry, host functions, lucinid gill microbiome composition, and gill microbiome functions are lacking. Similar approaches have been applied to study symbiosis in other marine organisms (Carney *et al.*, 2007; Boutet *et al.*, 2011; Sanders *et al.*, 2013; Roder *et al.*, 2015) and have the potential to advance our current understanding of lucinid-symbiont-environment interactions. Currently, ecosystem components and functions linking lucinid bivalves, their gill microbiomes, and their habitats remain under-sampled. As such, intra-population, inter-population, and inter-host species variations of host functions and lucinid gill microbiomes remain poorly understood. Additionally, contextual analyses of spatial and geochemical influences on gill microbiome diversity and/or host-symbiont gene expression have only been performed on a limited number of lucinid species (Chapter III).

In this study, we focused on characterizing the gill microbiome composition, host-microbiome functions, and possible lucinid-symbiont-environment connections within a *Stewartia floridana* (Conrad, 1833) population inhabiting seagrass beds at Bokeelia Fishing Pier, Pine Island, Florida, USA. *Stewartia floridana* has been associated with the oxygen-releasing roots of the seagrass species, *Thalassia testudinum* and *Ruppia maritima*, in sulfide-rich sediments (Fisher and Hand, 1984) along St. Joseph Bay, Florida, north of our sampling site. For this investigation, we first sought to determine the consistency of the *S. floridana* gill microbiomes across specimens and investigate whether the gill microbiomes consist of different strains and species of thioautotrophic symbionts and/or

other bacterial taxa, as previously reported in *C. orbiculata* (Chapter III) and *P. pectinatus* (Chapter II). Next, we compared gill microbiome functions in *S. floridana* with clade A thioautotrophic lucinid symbiont functions in seagrass-dwelling lucinid species *Codakia orbicularis* (König *et al.*, 2016), *Ctena orbiculata* (Chapter III) and *Loripes orbiculatus* (Petersen *et al.*, 2016), as well as gill microbiome functions of the clade C thioautotrophic symbiont, *Kistimonas*-like species, and/or *Spirochaeta*-like species in the mangrove-dwelling lucinid species *Phacoides pectinatus* (Chapter II). Finally, to explore the potentially positive effects of seagrass coverage on host and microbiome gene expression, we compared metatranscriptomic profiles of *S. floridana* specimens in three separate quadrats covered with 100% *Halodule wrightii* (seagrass), 100% *Syringodium filiforme* (seagrass), and 80%-20% mix of sand and *Halodule wrightii*.

Materials and methods

Sample collection

Sampling was conducted in 2014, pursuant to the Florida Fish and Wildlife Conservation Commission Special Activity License (SAL-14-1599SR). *Stewartia floridana* samples were collected from 30 cm diameter quadrats along three 50 m transects at Bokeelia Fishing Pier, Pine Island, Florida, beginning on 31 July 2014 (Goemann, 2015; Long, 2016). Geochemistry measurement methods are described in Goemann (2015) and Long (2016). Foot and gill dissection and fixation were performed within 30 minutes of collection in absolute ethanol (specimens labelled alphabetically DDG through NNG were

used for amplicon sequencing and specimens S26, S27, S28, S29, and S44 were used for metagenomic sequencing) or RNAlater (the remaining specimens) (Chapter II). Procedures in Chapter II were used for nucleic acids extraction and quantification and cDNA synthesis.

Sequencing

DNA and cDNA extracted from four *S. floridana* gill specimens and DNA from one *S. floridana* foot specimen was submitted to Duke Center for Genomic and Computational Biology (Durham, NC, USA) for Illumina MiSeq 2x250 bp sequencing using V2 chemistry (San Diego, CA, USA). 16S rRNA gene libraries from 21 other *S. floridana* gill tissues were prepared and sequenced at Clemson University (SC, USA) using methods in Chapter II. DNA extracted from one *S. floridana* gill sample was fragmented with NEBNext® dsDNA Fragmentase (New England Biolabs, Ipswich, MA, USA) and used for library preparation with NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs). This library was outsourced for Illumina HiSeq 2500 2x125 bp sequencing by Clemson University Genomics Institute (CUGI; Clemson, SC, USA). Molecular Research LP (Shallowater, TX, USA) prepared libraries for five other *S. floridana* gill samples using 20 uL of DNA (≤ 2.5 ng/μL total concentration) with the Nextera DNA Sample Preparation Kit (Illumina) and sequenced the libraries with the Illumina HiSeq 2500 2x100 bp paired-end platform. RNA extracted from eight *S. floridana* gill specimens was prepared for metatranscriptomic sequencing on Duke Center for Genomic and Computational Biology's HiSeq 4000 2x150bp platform, as detailed in

Chapter II. Concentration and insert size determination for all sequenced libraries were performed using methods described in Chapter III.

Data analysis

Procedures described in Chapter II were used for processing 16S rRNA gene reads in Mothur v1.40.5 (Schloss *et al.*, 2009). The OTU table was sub-sampled to 2,280 (the smallest four-digit number) sequences, which eliminated one cDNA sample. Phylogenetic analysis of 16S rRNA gene sequences (K2+G model) was performed as described in Chapter II. Metagenomic and metagenomic reads were trimmed using methods in Chapter II. Each sequenced metagenomic library was assembled and binned individually and reads combined from all six libraries were also co-assembled and binned based on procedures in Chapter II. MAG read mapping, quality assessment, annotation, AAI and ANI calculation methods were as described in Chapter II. Phylogenomic analyses was performed using 23 single-copy genes present in all gammaproteobacterial genomes and four genes present in all Spirochaetia genomes (Table 4.1), according to methods in Chapter II. *De novo* metatranscriptomic read assembly, transcript-to-MAG mapping, transcript cluster (gene) quantification, cross-sample count normalization, and transcript annotation were performed using procedures in Chapter II. Differential gene analyses ($p<0.05$, >2 fold-change) were performed separately on three batch-removed gene expression matrices of transcript clusters individually mapped to nucleotide sequences from the thioautotrophic symbiont MAGs, *Spirochaeta*-like MAG, and phylum Mollusca, using methods described in Chapter III.

qPCR

16S rRNA gene fragments cloned from two *S. floridana* gill specimens showed 100% identity in the V4 region to *Ca.* Thiodiazotropha-like OTU 1 and to qPCR primers 1417F and 1580R that target the autotrophic symbiont in *P. pectinatus* (Chapter II). The V4 sequence of *Spirochaeta*-like OTU2 were also identical to qPCR primers 15F and 226R that target the *Spirochaeta*-like species in *P. pectinatus* (Chapter II). As such, both sets of primers were used to amplify OTU1 and OTU2, according to the protocol detailed in Chapter II. Cloning, Sanger sequencing of cloned inserts, preparation of absolute quantification standards, qPCR, and data analyses steps were performed as described in Chapter II.

Availability of data and materials

All specimens are cataloged at the South Dakota School of Mines and Technology, Museum of Geology, with details provided through the iDigBio portal (https://www.idigbio.org/portal/recordsets/db3181c9-48dd-489f-96ab-a5888f5a938c).

Sequence data are deposited at the National Center for Biotechnology Information (NCBI; NCBI Resource Coordinators, 2016) under the BioProject ID PRJNA451498. Accession numbers are listed in **Table 4.2**.

Table 4.1. Single-copy genes and their protein substitution models used in phylogenomic analyses.

Reference genomes	Gene	Protein substitution model
Gammaproteobacteria	*dnaG*	LGF
	frr	LGF
	infC	LGF
	nusA	LGF
	pgk	LGF
	rplC	LG
	rplD	LGF
	rplE	LGF
	rplF	LG
	rplM	LG
	rplN	LGF
	rplP	LG
	rpmA	RTREVF
	rpsB	LGF
	rpsC	LGF
	rpsE	RTREVF
	rpsI	DAYHOFFF
	rpsJ	LGF
	rpsK	RTREVF
	rpsM	DAYHOFFF
	rpsS	RTREVF
	smpB	LG
	tsf	LGF
Spirochaetia	*pgk*	LGF
	rplA	LGF
	rplK	LGF
	rpmA	WAGF

Table 4.2. NCBI accession numbers of raw read and sequence data generated in this chapter. All data are linked to NCBI's BioProject ID PRJNA451498.

Database	Accession numbers	Dataset description
Sequence Read Archive (SRA)	SRX3040972-SRX3040986; SRX3040997-SRX3041004; SRX3041010- SRX3041014	Amplicon-sequenced read data (V4 region of 16S rRNA gene)
	SRX3040871-SRX3040876 SRX3040972-SRX3040974	Metagenomic read data
	SRR7949662-SRR7949669	Metatranscriptomic read data
GenBank	MH999890- MH999899	Sequences of top ten most abundant OTUs
	MF974564	Sequence of *Spirochaeta*-like OTU 281
	MF973039	Full-length 16S rRNA gene sequence from MAG of Spirochaetia sp.
Genome	NATX00000000-NAUC00000000	MAGs of thioautotrophic symbionts
	NAUD00000000	MAG of Spirochaetia sp.

Results

Site characterization

Stewartia floridana abundances, vegetation cover, and geochemistry of the study site at Bokeelia Pier, Pine Island (Figure 4.1) were already assessed and described in detail (Goemann, 2015; Long, 2016). Briefly, live *S. floridana* specimens were observed in quadrats with 0%-100% coverage of *Halodule wrightii*, *Syringodium filiforme*, and/or *Thalassia testudinum* (Figure 4.1), and their abundances correlated positively with ranked percentages of seagrass coverage (Goemann, 2015). Previous analysis revealed no statistically significant differences in measured porewater geochemistry parameters across quadrats with 100% sand coverage, 100% seagrass coverage and mixed sand-seagrass coverage (Long, 2016).

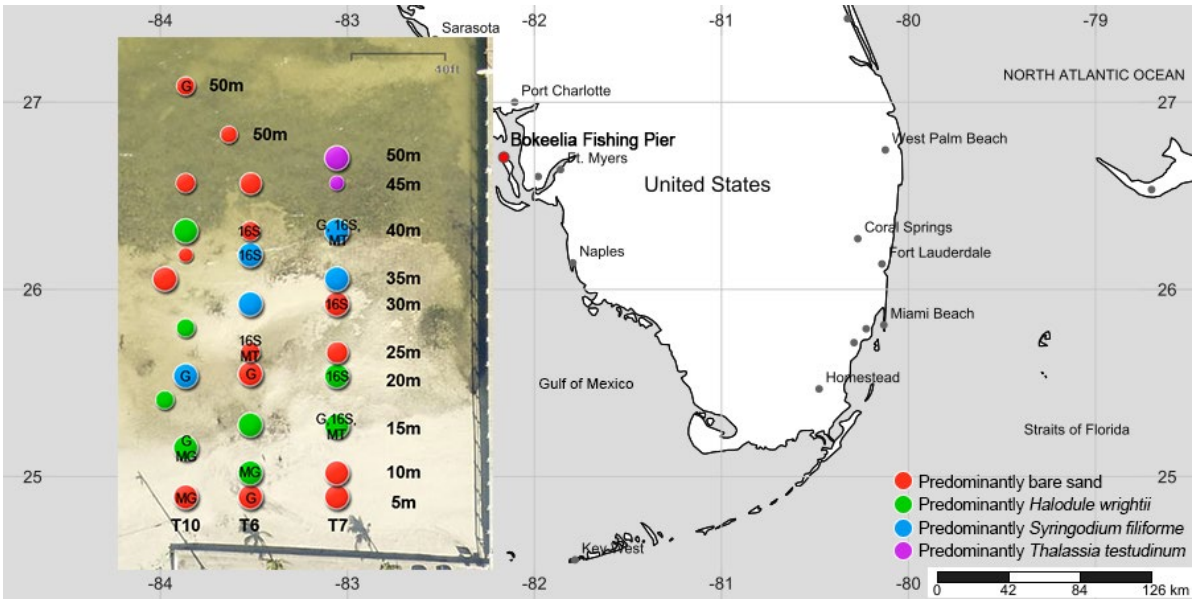
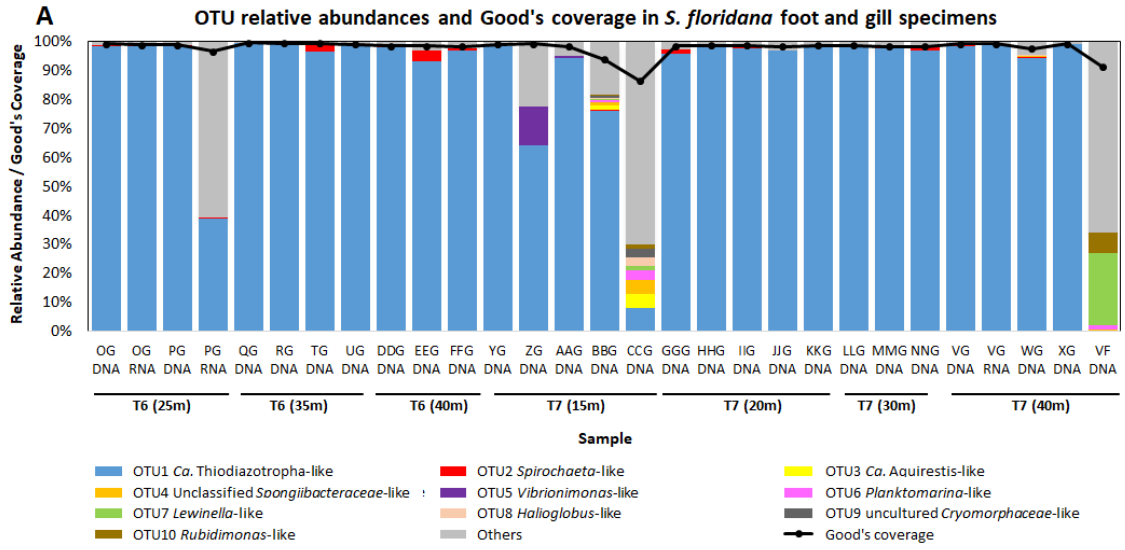


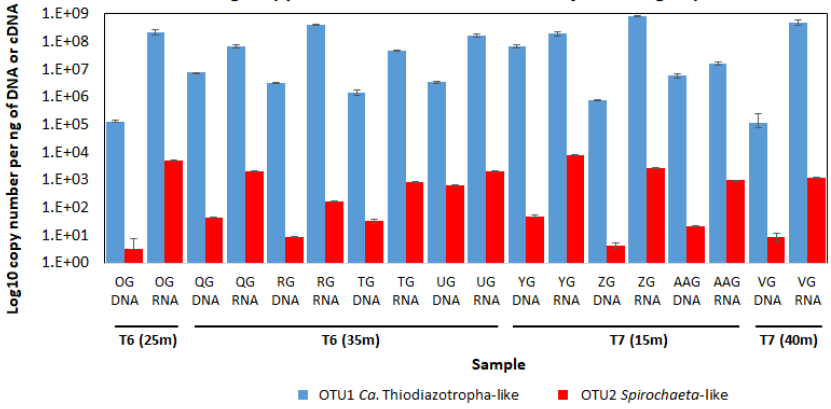
Figure 4.1. Map showing location of the sampling site at Bokeelia Pier, Florida, USA and sampling scheme (insert; satellite data: © Esri). Quadrats in insert were identified by transect IDs (x-axis) and distance to shore (y-axis). Quadrats (circles) labelled with "G" were characterized with geochemical methods, while quadrats where gill specimens were collected sequencing were labelled with "16S" (amplicon sequencing), "MG" (metagenomic sequencing) and/or "MG" (metatranscriptomic sequencing). The size of each circle is proportionate to the percentage of sand or vegetation coverage of the quadrat.

Gill microbiome diversity

The 16S rRNA gene V4 region in 25 *S. floridana* gill DNA samples, three gill cDNA samples, and one foot specimen (DNA) were sequenced and resulted in 87-100% Good's coverage values (Good, 1953) (Figure 4.2). All but two gill DNA and cDNA samples consisted of >50% relative abundances of a *Ca.* Thiodiazotropha-like operational taxonomic unit (OTU) affiliated with clade A thioautotrophic lucinid symbionts (Figures 4.2 and 4.3). Metagenomic sequencing of six *S. floridana* gill specimens generated gammaproteobacterial MAGs with 16S rRNA gene V4 sequences identical to the predominant *Ca.* Thiodiazotropha-like OTU1 (Table 4.3). These MAGs shared ≥97% pairwise average nucleotide identity (pANI), ≥98 pairwise average amino acid identity (pAAI), and 74±13% orthologous fraction (OF) with each other. The MAGs were most closely related and shared 88±1% pANI, 93±0.3% pAAI, and 77±10% OF to the *Loripes orbiculatus* symbiont, *Ca.* Thiodiazotropha endoloripes (Petersen *et al.*, 2016), also belonging to clade A (Figure 4.3b). Based on the 85-90% pAAI species boundaries defined

in Rodriguez-R and Konstantinidis (2014), the thioautotrophic *S. floridana* symbiont likely belongs to the same species as *Ca.* Thiodiazotropha endoloripes. The <90% ANI and <95% AAI values shared between the *S. floridana* symbiont and *Ca.* Thiodiazotropha endoloripes suggest that both could be different strains of the same species. Because the existing name of this symbiont species, *Ca.* Thiodiazotropha endoloripes, does not accurately reflect the range of lucinid host species this symbiont can inhabit, we propose an amendment of the species name from *Ca.* Thiodiazotropha endoloripes to *Ca.* Thiodiazotropha endolucininae, where "endo" ("within") refers to the intracellular location of the gill symbiont and "Lucininae" refers to members of the Lucininae clam subfamily, including *Loripes orbiculatus* and *Stewartia floridana*, which, to date, are exclusively associated with this symbiont species. *Candidatus* Thiodiazotropha endolucininae spp. shared 70±2% pANI, 90±0% pAAI, and 33±0% OF with the free-living gammaproteobacterial species *Sedimenticola selenatireducens* DSM 17993 that inhabits estuarine sediments (Benson *et al.*, 2014; unpublished GenBank accession ATZE01000000). Members of *Ca.* Thiodiazotropha endolucininae formed a sister group to another clade comprising other clade A symbionts, including *Ca.* Thiodiazotropha endolucinida and *Ca.* Thiodiazotropha endolucinidaduo associated with *Codakia orbicularis* (*Ca.* Thiodiazotropha endolucinida; König *et al.*, 2016) and *Ctena orbiculata* symbionts (both species; Chapter III; Figure 4.3b). Average pANI, pAAI, and OF values shared between MAGs of *Ca.* Thiodiazotropha endolucininae and the clade comprising *Ca.* Thiodiazotropha endolucinida and *Ca.* Thiodiazotropha endolucinidaduo were 73±4%, 74±0.2%, and 63±7%, respectively.


A
OTU relative abundances and Good's coverage in S. floridana foot and gill specimens
Relative Abundance / Good's Coverage
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
OG DNA
OG RNA
PG DNA
PG RNA
QG DNA
RG DNA
TG DNA
UG DNA
DDG DNA
EEG DNA
FFG DNA
YG DNA
ZG DNA
AAG DNA
BBG DNA
CCG DNA
GGG DNA
HHG DNA
IIG DNA
JJG DNA
KKG DNA
LLG DNA
MMG DNA
NNG DNA
VG DNA
VG RNA
WG DNA
XG DNA
VF DNA
T6 (25m)
T6 (35m)
T6 (40m)
T7 (15m)
T7 (20m)
T7 (30m)
T7 (40m)
Sample
OTU1 Ca. Thiodiazotropha-like
OTU2 Spirochaeta-like
OTU3 Ca. Aquirestis-like
OTU4 Unclassified Spongiibacteraceae-like
OTU5 Vibrionimonas-like
OTU6 Planktomarina-like
OTU7 Lewinella-like
OTU8 Halioglobus-like
OTU9 uncultured Cryomorphaceae-like
OTU10 Rubidimonas-like
Others
Good's coverage


B
Average copy numbers of OTU1 and OTU2 in S. floridana gill specimens
Log10 copy number per ng of DNA or cDNA
1.E+09
1.E+08
1.E+07
1.E+06
1.E+05
1.E+04
1.E+03
1.E+02
1.E+01
1.E+00
OG DNA
OG RNA
QG DNA
QG RNA
RG DNA
RG RNA
TG DNA
TG RNA
UG DNA
UG RNA
YG DNA
YG RNA
ZG DNA
ZG RNA
AAG DNA
AAG RNA
VG DNA
VG RNA
T6 (25m)
T6 (35m)
T7 (15m)
T7 (40m)
Sample
OTU1 Ca. Thiodiazotropha-like
OTU2 Spirochaeta-like

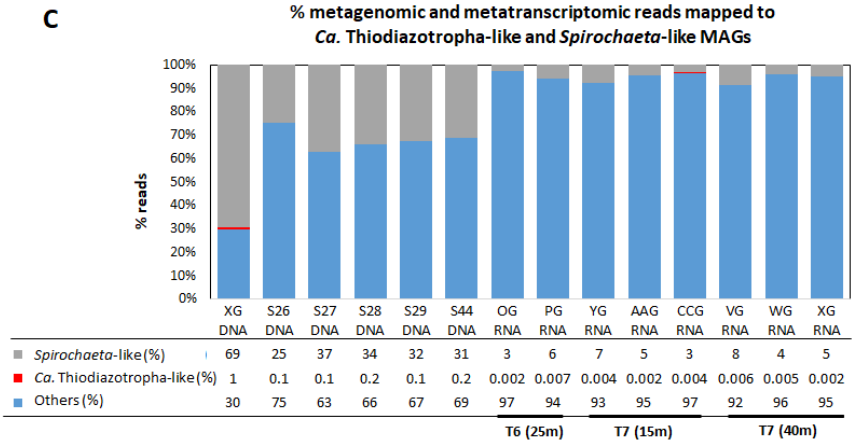

C
% metagenomic and metatranscriptomic reads mapped to Ca. Thiodiazotropha-like and Spirochaeta-like MAGs
% reads
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
XG DNA
S26 DNA
S27 DNA
S28 DNA
S29 DNA
S44 DNA
OG RNA
PG RNA
YG RNA
AAG RNA
CCG RNA
VG RNA
WG RNA
XG RNA
Spirochaeta-like (%) 69 25 37 34 32 31 3 6 7 5 3 8 4 5
Ca. Thiodiazotropha-like (%) 1 0.1 0.1 0.2 0.1 0.2 0.002 0.007 0.004 0.002 0.004 0.006 0.005 0.002
Others (%) 30 75 63 66 67 69 97 94 93 95 97 92 96 95
T6 (25m)
T7 (15m)
T7 (40m)

Figure 4.2. (A) Relative abundances and Good's coverages of subsampled bacterial OTUs in *S. floridana* gill DNA, gill cDNA and foot DNA (VF) samples. (B) qPCR copy numbers and standard error bars of *Ca.* Thiodiazotropha-like OTU1 and *Spirochaeta*-like OTU2 in *S. floridana* gill DNA and cDNA samples. (C) Percentages of metagenomic and metatranscriptomic reads mapped to the representative *Ca.* Thiodiazotropha lucininae MAG (S27) and the *Spirochaeta*-like MAG.

Besides the thioautotrophic symbiont, 332 *Spirochaeta*-like OTUs were also predicted in *S. floridana* gill specimens. The most abundant of these, OTU2, occurred in 18 of 25 amplicon sequenced gill specimens at average 0.7±1% relative abundance and in all three cDNA samples at average 0.3±0.2% relative abundance (Figure 4.2a). From the *S. floridana* gill co-assembly, we binned a low-quality ~22% complete *Spirochaeta*-like MAG without a 16S rRNA gene sequence and recovered a separate 7,016 bp unbinned contig 91% identical to OTU2 and 100% identical to another *Spirochaeta*-like OTU 281 occurring at 0.0004 relative abundance in gill specimen CCG in the V4 region. From the gill metatranscriptomes, we also identified a transcript 99% identical to OTU2 within a transcript cluster expressed at average 0.8±0.7 trimmed mean of M-values normalized transcripts per million (TPM).

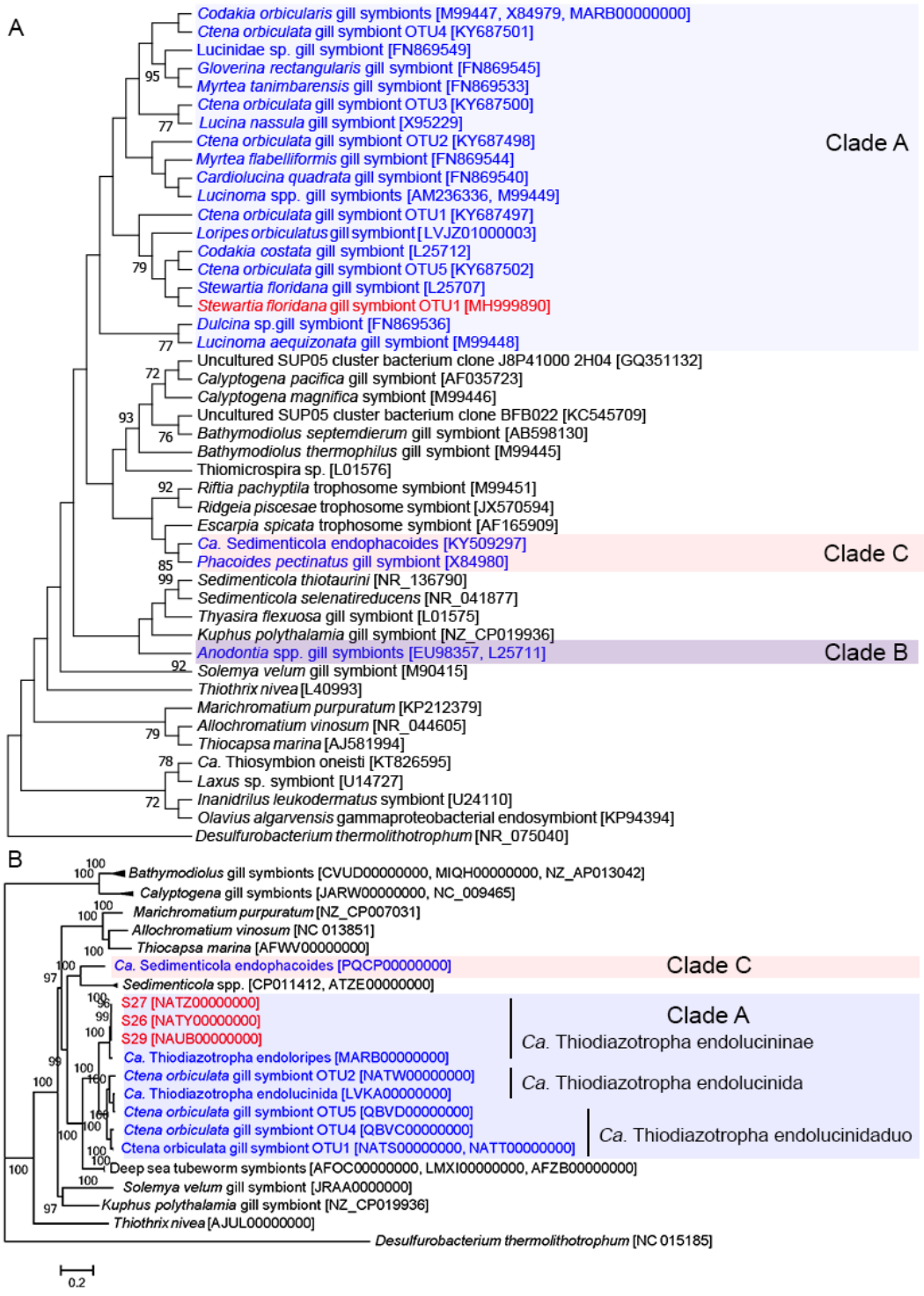

A
Codakia orbicularis gill symbionts [M99447, X84979, MARB00000000]
Ctena orbiculata gill symbiont OTU4 [KY687501]
Lucinidae sp. gill symbiont [FN869549]
Gloverina rectangularis gill symbiont [FN869545]
Myrtea tanimbarensis gill symbiont [FN869533]
Ctena orbiculata gill symbiont OTU3 [KY687500]
Lucina nassula gill symbiont [X95229]
Ctena orbiculata gill symbiont OTU2 [KY687498]
Myrtea flabelliformis gill symbiont [FN869544]
Cardiolucina quadrata gill symbiont [FN869540]
Lucinoma spp. gill symbionts [AM236336, M99449]
Ctena orbiculata gill symbiont OTU1 [KY687497]
Loripes orbiculatus gill symbiont [LVJZ01000003]
Codakia costata gill symbiont [L25712]
Ctena orbiculata gill symbiont OTU5 [KY687502]
Stewartia floridana gill symbiont [L25707]
Stewartia floridana gill symbiont OTU1 [MH999890]
Dulcina sp.gill symbiont [FN869536]
Lucinoma aequizonata gill symbiont [M99448]
Clade A
Uncultured SUP05 cluster bacterium clone J8P41000 2H04 [GQ351132]
Calyptogena pacifica gill symbiont [AF035723]
Calyptogena magnifica symbiont [M99446]
Uncultured SUP05 cluster bacterium clone BFB022 [KC545709]
Bathymodiolus septemdierum gill symbiont [AB598130]
Bathymodiolus thermophilus gill symbiont [M99445]
Thiomicrospira sp. [L01576]
Riftia pachyptila trophosome symbiont [M99451]
Ridgeia piscesae trophosome symbiont [JX570594]
Escarpia spicata trophosome symbiont [AF165909]
Ca. Sedimenticola endophacoides [KY509297]
Phacoides pectinatus gill symbiont [X84980]
Clade C
Sedimenticola thiotaurini [NR_136790]
Sedimenticola selenatireducens [NR_041877]
Thyasira flexuosa gill symbiont [L01575]
Kuphus polythalamia gill symbiont [NZ_CP019936]
Anodontia spp. gill symbionts [EU98357, L25711]
Clade B
Solemya velum gill symbiont [M90415]
Thiothrix nivea [L40993]
Marichromatium purpuratum [KP212379]
Allochromatium vinosum [NR_044605]
Thiocapsa marina [AJ581994]
Ca. Thiosymbion oneisti [KT826595]
Laxus sp. symbiont [U14727]
Inanidrilus leukodermatus symbiont [U24110]
Olavius algarvensis gammaproteobacterial endosymbiont [KP94394]
Desulfurobacterium thermolithotrophum [NR_075040]
B
Bathymodiolus gill symbionts [CVUD00000000, MIQH00000000, NZ_AP013042]
Calyptogena gill symbionts [JARW00000000, NC_009465]
Marichromatium purpuratum [NZ_CP007031]
Allochromatium vinosum [NC 013851]
Thiocapsa marina [AFWV00000000]
Ca. Sedimenticola endophacoides [PQCP00000000]
Clade C
Sedimenticola spp. [CP011412, ATZE00000000]
S27 [NATZ00000000]
S26 [NATY00000000]
S29 [NAUB00000000]
Clade A
Ca. Thiodiazotropha endolucininae
Ca. Thiodiazotropha endoloripes [MARB00000000]
Ctena orbiculata gill symbiont OTU2 [NATW00000000]
Ca. Thiodiazotropha endolucinida [LVKA00000000]
Ca. Thiodiazotropha endolucinida
Ctena orbiculata gill symbiont OTU5 [QBVD00000000]
Ctena orbiculata gill symbiont OTU4 [QBVC00000000]
Ctena orbiculata gill symbiont OTU1 [NATS00000000, NATT00000000]
Ca. Thiodiazotropha endolucinidaduo
Deep sea tubeworm symbionts [AFOC00000000, LMXI00000000, AFZB00000000]
Solemya velum gill symbiont [JRAA0000000]
Kuphus polythalamia gill symbiont [NZ_CP019936]
Thiothrix nivea [AJUL00000000]
Desulfurobacterium thermolithotrophum [NC 015185]
0.2

Figure 4.3. (A) Bootstrap consensus maximum likelihood tree based on the 16S rRNA gene sequence and (B) phylogenomic tree based on 23 marker genes sequenced from *Ca.* Thiodiazotropha endolucininae in *S. floridana* (red text), in relation to thioautotrophic lucinid symbionts (blue text), marine symbionts and gammaproteobacterial free-living species. GenBank (Benson *et al.*, 2014) accession numbers are indicated in square brackets. The outgroup used in both trees was *Desulfurobacterium thermolithotrophum* from phylum Aquificae. Tree nodes show bootstrap values of >70% (A) and approximate likelihood-ratio test (aLRT) SH-like support values (B; Anisimova and Gascuel, 2006). The scale bar in B indicates 0.2 substitution per site.

Table 4.3. General features of metagenome-assembled genomes (MAGs) from *S. floridana* gill specimens. Abbreviation: PEG, protein-encoding genes.

Categorized species	MAG ID	Transect/ quadrat	Size	Contigs	PEGs	G+C	N50	Completeness		Contamination	Strain hetero-geneity	MAG quality[c]
			Mb	#	#	%	Kb	CheckM[a]	BUSCO[b]	%	%	
Ca. Thiodiazotropha endolucininae	XG	T7/40m	1.6	346	1,420	54	5	50	33	4	100	Low
	S26	T10/5m	4.2	161	3,773	53	38	96	92	1	14	High
	S27	T10/10m	4.5	58	4,018	53	99	99	95	1	0	High
	S28	T10/10m	2.9	535	2,639	54	7	64	62	0	0	Medium
	S29	T10/10m	4.6	36	4,028	53	24	99	96	1	0	High
	S44	T6/10m	3.3	822	2,964	53	5	82	65	2	27	Medium
Spirochaeta-like sp.	NA (co-assembly	NA (co-assembly)	0.8	232	808	53	4	26	18	0	0	Low

[a]Parks *et al.*, 2015; [b]Simao *et al.*, 2015; [c]Bowers *et al.*, 2017

On the 16S rRNA gene tree, sequences of *Spirochaeta*-like OTU2, OTU281, and related sequences in the gill metatranscriptomes and metagenomes were placed in a distinct monophyletic clade together with the *Spirochaeta*-like species in the *P. pectinatus* gill microbiome (Chapter II) at 91-95% identity in the V4 region, 69% pANI, 67% pAAI, and 58% OF, as well as with proposed spirochete endosymbionts in *Olavius* gutless marine worms (Blazejak *et al.*, 2005; Dubilier *et al.*, 1999) at 90-94% identity in the V4 region (Figure 4.4a). Their common closest free-living 16S rRNA gene relatives include *Spirochaeta aurantia* from an Austrian freshwater lake at 81-89% identity in the V4 region (Hahn *et al.*, 2004) and *S. halophila* from the black mud of an Egyptian saline solar lake at 83% identity in the V4 region (ATCC® 29478™; Figure 4.4a). The closest free-living genomic relative to the *Spirochaeta*-like MAG was likely *Spirochaeta thermophila* (no assignable pANI value, 50% pAAI, and 69% OF) (Figure 4.4b). qPCR assays successfully amplified sequences of *Ca. Thiodiazotropha*-like OTU1 and *Spirochaeta*-like OTU2 in a subset of matched cDNA and DNA samples from amplicon-sequenced gill specimens (Figure 4.2b). There were 13±25 million (38±16%) paired-end metagenomic reads and 2±2 million (5±2%) paired-end metatranscriptomic reads mapped to the representative *Ca.* Thiodiazotropha-like MAG S27, whereas 0.2±0.4 million (0.3±0.4%) metagenomic reads and 1,434±959 (0.004±0.002%) transcriptomic reads mapped to the *Spirochaeta*-like MAG (Figure 4.2c).

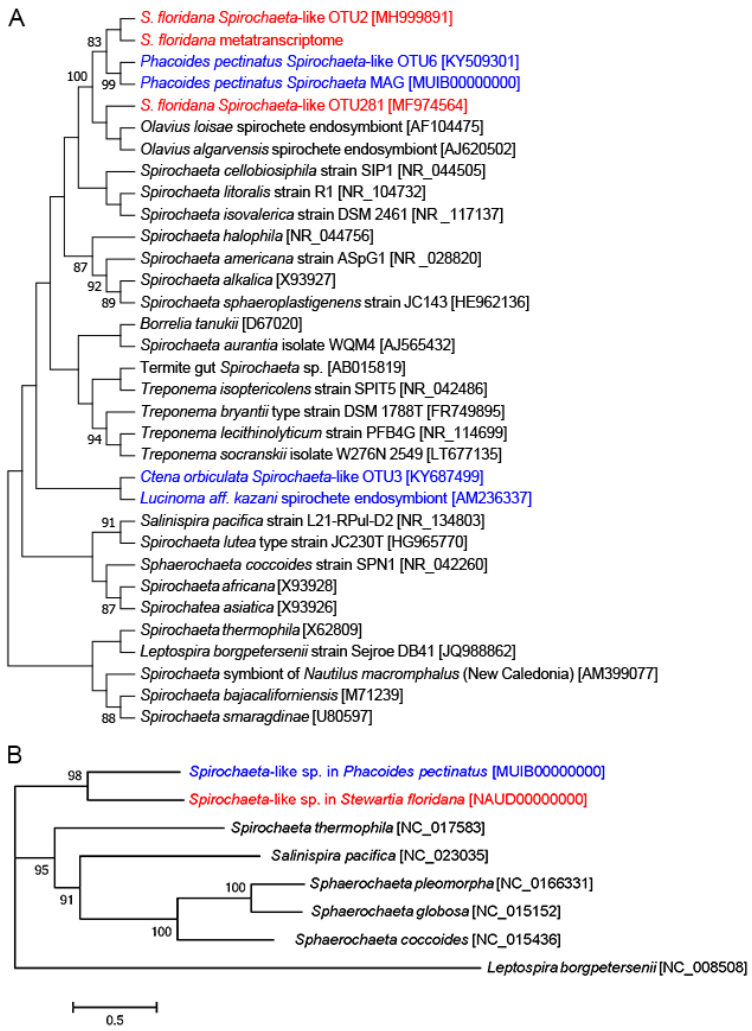


Figure 4.4. Unrooted (A) bootstrap consensus maximum likelihood tree based on 16S rRNA gene sequences and (B) phylogenomic tree based on four marker genes from *Spirochaeta*-like species sequenced in *S. floridana* (red text), in relation to spirochete species associated with lucinid clams (blue text), spirochete symbionts in marine species and free-living species. GenBank (Benson *et al.*, 2014) accession numbers are indicated in square brackets. Tree nodes show bootstrap values of >70% (A) and approximate likelihood-ratio test (aLRT) SH-like support values (B; Anisimova and Gascuel, 2006). The scale bar in B indicates 0.5 substitution per site.

Thioautotrophic symbiont functions

The 35 most abundantly expressed protein-coding transcript clusters mapped to MAGs of *Ca.* Thiodiazotropha endolucininae associated with *S. floridana* included heat shock (average 499±513 TPM), transport (average 151±233 TPM), transferase (average 69±80 TPM), transcriptional regulation (average 57±99 TPM), and protein degradation (average 18±9 TPM) functions, among others (Figure 4.5a). Growth-related genes, such as those encoding DNA-directed RNA polymerase subunit alpha/50S ribosomal protein L4 (average 7±9 TPM) and RNA polymerase factor sigma-54 (average 6±4 TPM), were also expressed by the symbiont. Thioautotrophy-related genes encoding the dissimilatory sulfite reductase alpha subunit (DsrA; average 57±49 TPM), sulfurtransferases (average 15±10 TPM), nickel-dependent hydrogenase large subunit (average 14±13 TPM), large chain form IAq ribulose bisphosphate carboxylase (RuBisCO; average 34±40 TPM), and fructose-bisphosphate aldolase (average 27±25 TPM) were also among the most abundantly expressed in the bacterial transcriptomes (Figure 4.5a). *Candidatus* Thiodiazotropha endolucininae can potentially derive energy through sulfur oxidation (average 14±25 TPM), hydrogen oxidation (average 3±7 TPM), and C1-compound oxidation (average 1±1 TPM) in *S. floridana* (Figure 4.6a). Energy obtained is likely used for carbon (average 21±31 TPM) and nitrogen fixation (average 1±2 TPM; Figures 4.7-4.8). Besides autotrophy, mixotrophy-related genes participating in the pentose phosphate pathway (average 5±5TPM), glycolysis (average 16±19 TPM), TCA cycle (average 3±3 TPM), and organic compound transport (average 7±6 TPM) were also identified in the bacterial MAGs and transcriptomes (Figure 4.7a). In *S. floridana*, *Ca.* Thiodiazotropha

endolucininae could potentially assimilate nitrate and nitrite (average 2±2 TPM), as well as urea (average 2±4 TPM) for nitrogen (Figure 4.8). The species is genetically capable of aerobic respiration with cbb3 and aa3 type terminal oxidases (average 5±8 TPM) and denitrification (average 3±4TPM) (Figure 4.8). *Candidatus* Thiodiazotropha endolucininae encoded and expressed genes related to type I (average 1±2 TPM), type II (average 4±5TPM), and type VI (average 4±5 TPM) secretion systems in *S. floridana* (Figure 4.6a). Biosynthesis genes for all twenty essential amino acids (average 5±8TPM), as well as vitamins B1 (average 3±2 TPM), B2 (average 2±2 TPM), B6 (average 3±2 TPM), B7 (average 3±2TPM), and B9 (average 3±2 TPM) were detected in the bacterial MAGs and transcriptomes (Figure 4.6b). The species also encoded and expressed genes involved in flagellar (average 3±4 TPM), chemotaxis (average 1±1 TPM), iron uptake (average 2±3 TPM), and phosphate uptake (average 2±3 TPM). *Candidatus* Thiodiazotropha endolucininae transcriptomes sequenced from *S. floridana* specimens collected from quadrats with different vegetation coverages did not show clear quadrat-specific clustering patterns on their gene expression count matrix based on principal component analysis (PCA; Figure 4.9a). Only four differentially expressed (DE; $p<0.05$, ≥ 2 fold-change) genes were predicted across these quadrats by voom (Law *et al.*, 2014). These included two proteins with domains of unknown functions, a chemotaxis protein (CheR)-encoding transcript (upregulated in quadrat with 100% *H. wrightiii* coverage compared to quadrat with 100% *Syringodium filiforme* coverage), and a glycine cleavage system aminomethyltransferase (GcvT)-encoding transcript (upregulated in quadrat with 100% *S. filiforme* compared to quadrat with 100% *H. wrightiii*) (Figure 4.10).

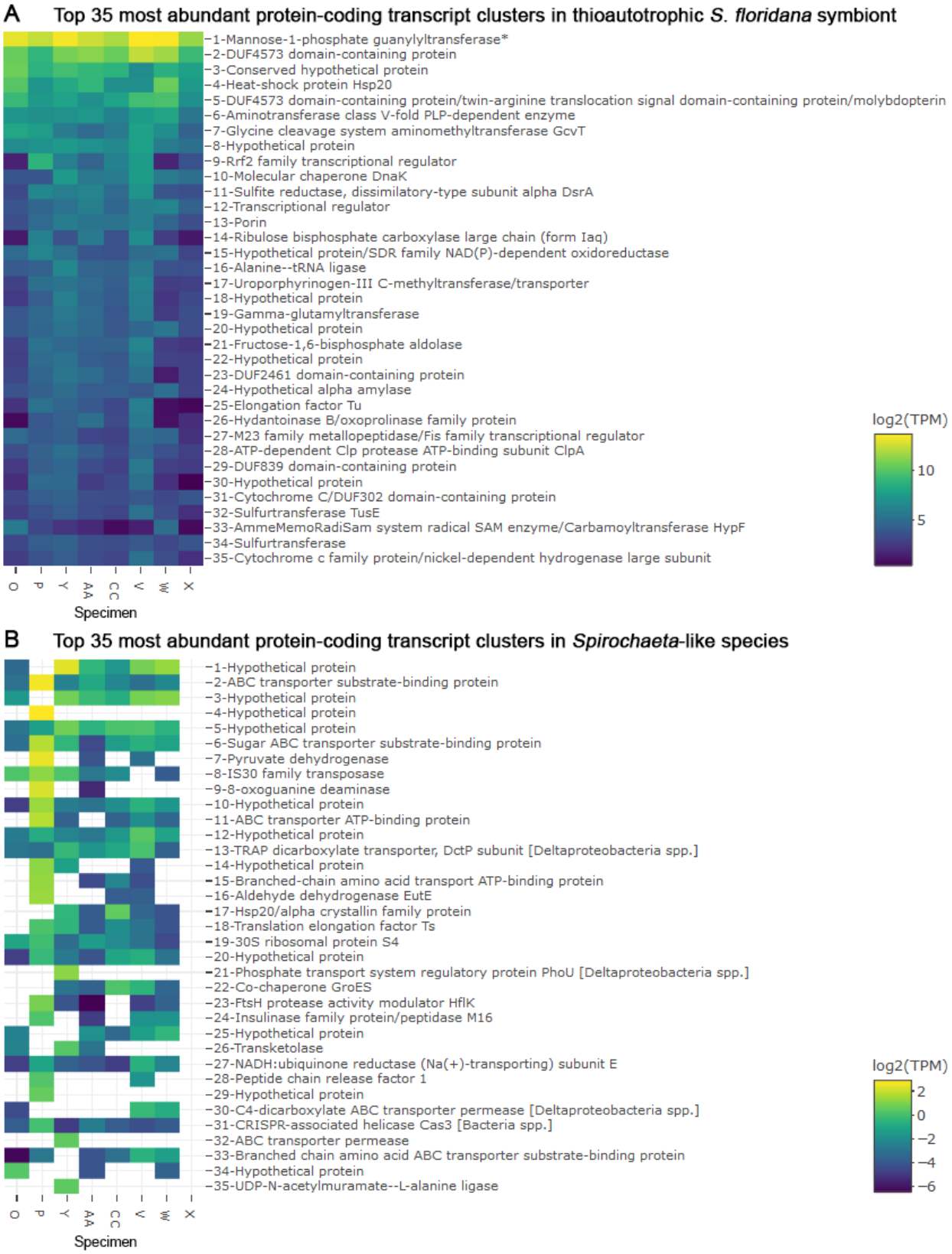
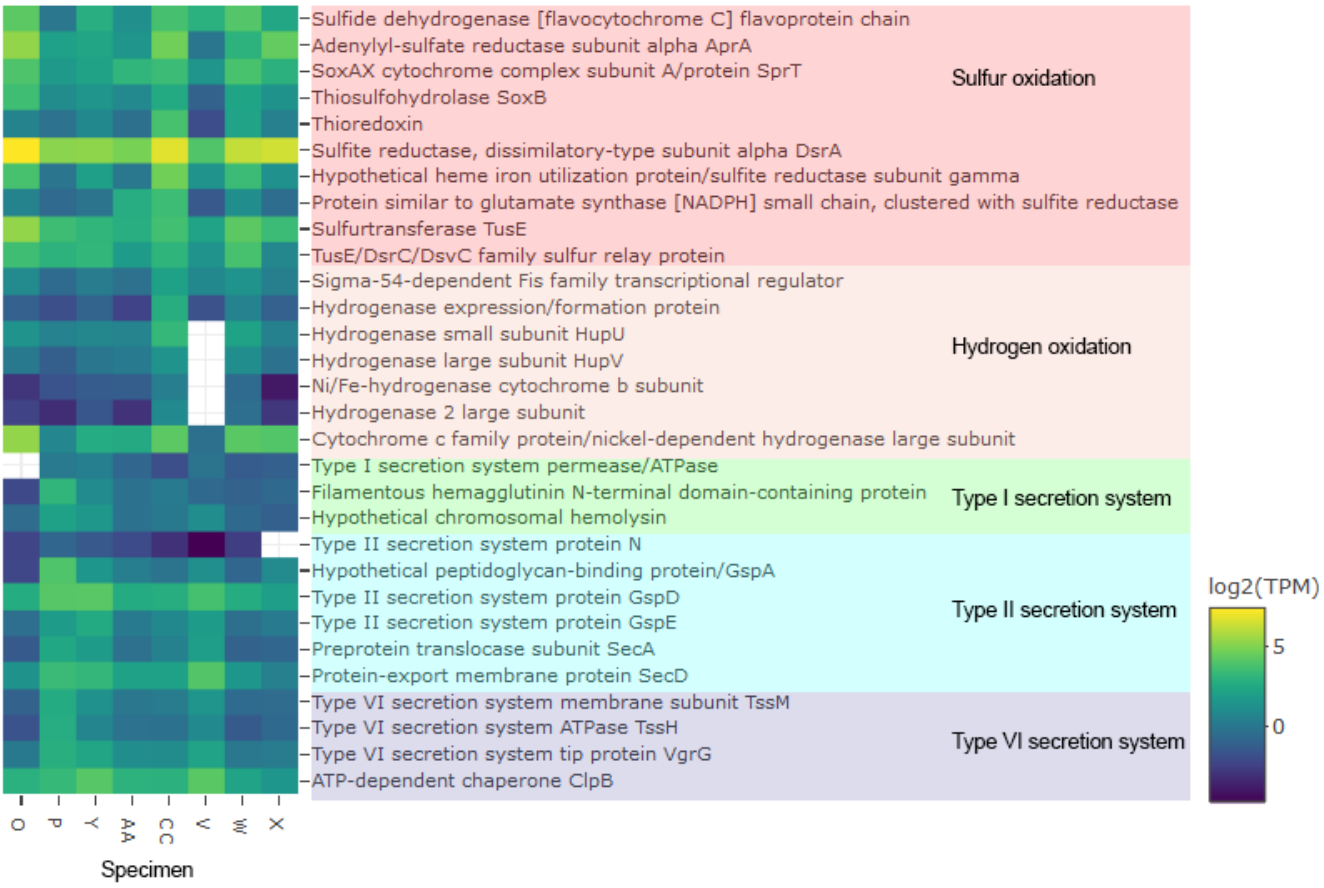

A
Top 35 most abundant protein-coding transcript clusters in thioautotrophic S. floridana symbiont
1-Mannose-1-phosphate guanylyltransferase*
2-DUF4573 domain-containing protein
3-Conserved hypothetical protein
4-Heat-shock protein Hsp20
5-DUF4573 domain-containing protein/twin-arginine translocation signal domain-containing protein/molybdopterin
6-Aminotransferase class V-fold PLP-dependent enzyme
7-Glycine cleavage system aminomethyltransferase GcvT
8-Hypothetical protein
9-Rrf2 family transcriptional regulator
10-Molecular chaperone DnaK
11-Sulfite reductase, dissimilatory-type subunit alpha DsrA
12-Transcriptional regulator
13-Porin
14-Ribulose bisphosphate carboxylase large chain (form Iaq)
15-Hypothetical protein/SDR family NAD(P)-dependent oxidoreductase
16-Alanine--tRNA ligase
17-Uroporphyrinogen-III C-methyltransferase/transporter
18-Hypothetical protein
19-Gamma-glutamyltransferase
20-Hypothetical protein
21-Fructose-1,6-bisphosphate aldolase
22-Hypothetical protein
23-DUF2461 domain-containing protein
24-Hypothetical alpha amylase
25-Elongation factor Tu
26-Hydantoinase B/oxoprolinase family protein
27-M23 family metallopeptidase/Fis family transcriptional regulator
28-ATP-dependent Clp protease ATP-binding subunit ClpA
29-DUF839 domain-containing protein
30-Hypothetical protein
31-Cytochrome C/DUF302 domain-containing protein
32-Sulfurtransferase TusE
33-AmmeMemoRadiSam system radical SAM enzyme/Carbamoyltransferase HypF
34-Sulfurtransferase
35-Cytochrome c family protein/nickel-dependent hydrogenase large subunit
log2(TPM)
10
5
O
P
Y
AA
CC
V
W
X
Specimen
B
Top 35 most abundant protein-coding transcript clusters in Spirochaeta-like species
1-Hypothetical protein
2-ABC transporter substrate-binding protein
3-Hypothetical protein
4-Hypothetical protein
5-Hypothetical protein
6-Sugar ABC transporter substrate-binding protein
7-Pyruvate dehydrogenase
8-IS30 family transposase
9-8-oxoguanine deaminase
10-Hypothetical protein
11-ABC transporter ATP-binding protein
12-Hypothetical protein
13-TRAP dicarboxylate transporter, DctP subunit [Deltaproteobacteria spp.]
14-Hypothetical protein
15-Branched-chain amino acid transport ATP-binding protein
16-Aldehyde dehydrogenase EutE
17-Hsp20/alpha crystallin family protein
18-Translation elongation factor Ts
19-30S ribosomal protein S4
20-Hypothetical protein
21-Phosphate transport system regulatory protein PhoU [Deltaproteobacteria spp.]
22-Co-chaperone GroES
23-FtsH protease activity modulator HflK
24-Insulinase family protein/peptidase M16
25-Hypothetical protein
26-Transketolase
27-NADH:ubiquinone reductase (Na(+)-transporting) subunit E
28-Peptide chain release factor 1
29-Hypothetical protein
30-C4-dicarboxylate ABC transporter permease [Deltaproteobacteria spp.]
31-CRISPR-associated helicase Cas3 [Bacteria spp.]
32-ABC transporter permease
33-Branched chain amino acid ABC transporter substrate-binding protein
34-Hypothetical protein
35-UDP-N-acetylmuramate--L-alanine ligase
log2(TPM)
2
0
−2
−4
−6
O
P
Y
AA
CC
V
W
X
Specimen

Figure 4.5. Log2-transformed TMM-normalized TPM of gene products of the 35 most abundantly expressed protein-coding transcript clusters mapped to MAGs of the (A) thioautotrophic *S. floridana* symbiont (*Ca.* Thiodiazotropha endolucininae) and (B) *Spirochaeta*-like species. Transcript clusters with zero TPM values are represented as white cells. '*' in (A) denotes a multi-mapping transcript cluster. Abbreviations: DUF, domain of unknown function; PLP, pyridoxal-5'-phosphate; SDR, short-chain dehydrogenases/reductases; Tus, sulfur carrier proteins homologous to some Dsr proteins; SAM, S-adenosyl-L-methionine; ABC, ATP-binding cassette; ATP; adenosine triphosphate; IS, insertion sequence; TRAP, tripartite ATP-independent periplasmic transporter; Dct, dicarboxylate transport proteins; Hsp, heat shock protein; FtsH, ATP-dependent zinc metalloprotease; NADH, reduced nicotinamide adenine dinucleotide; CRISPR, clustered regularly-interspaced short palindromic repeats; UDP, uridine diphosphate.

A Lithotrophy and secretion systems-related transcript clusters in thioautotrophic *S. floridana* symbiont



B Amino acids and B vitamins-related transcript clusters in thioautotrophic *S. floridana* symbiont

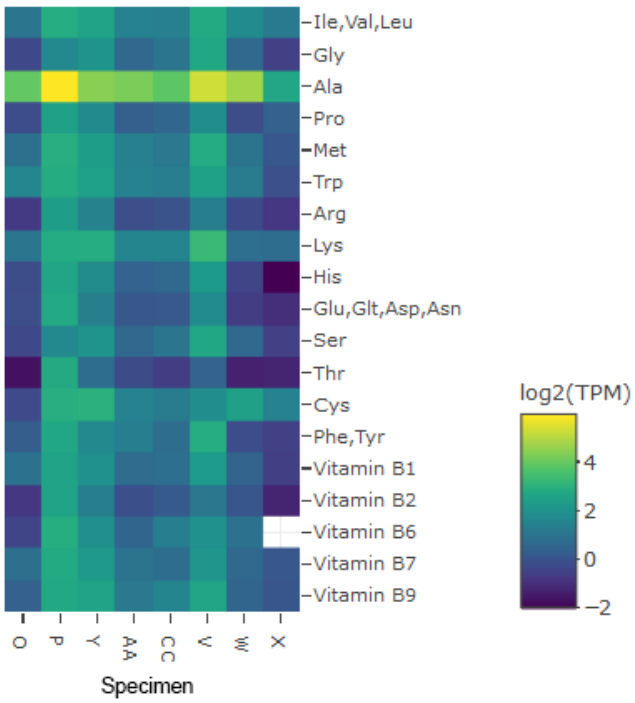


Figure 4.6. Log_2-transformed TMM-normalized TPM of gene products of (A) lithotrophy and secretion system-related and (B) amino acids and B vitamins-related transcript clusters mapped to the thioautotrophic *S. floridana* symbiont (*Ca.* Thiodiazotropha endolucininae). Transcript clusters with zero TPM values are represented as white cells. Abbreviations: Sox, sulfur oxidation enzyme; NADPH, reduced nicotinamide adenine dinucleotide phosphate; ATP, adenosine triphosphate; Gsp, general secretory pathway protein; Sec, secretory export protein; Tss, type VI secretion system proteins; VgrG, valine-glycine repeat protein G; ClpB, caseinolytic peptidase B protein homolog.

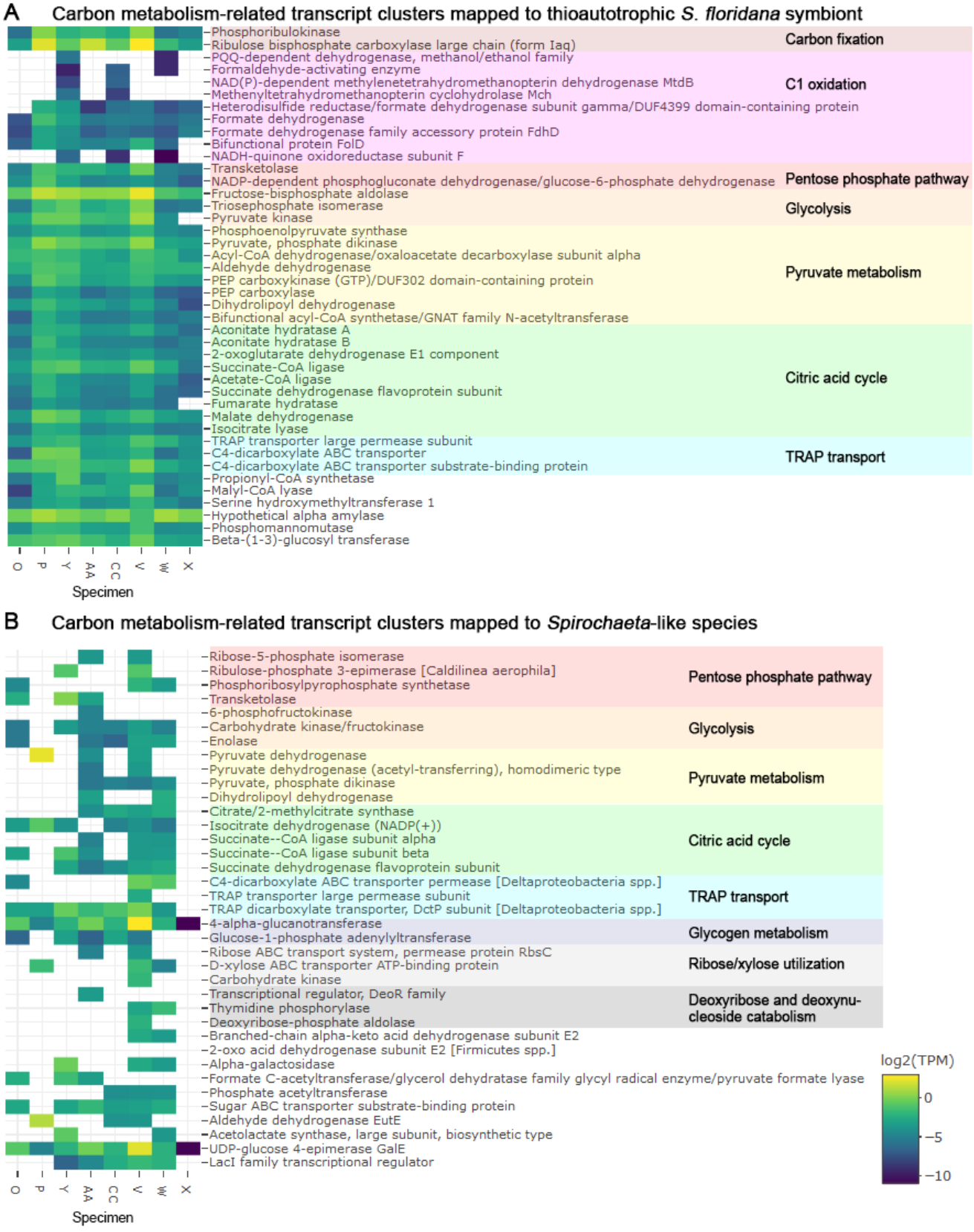

A
Carbon metabolism-related transcript clusters mapped to thioautotrophic S. floridana symbiont
Phosphoribulokinase
Ribulose bisphosphate carboxylase large chain (form Iaq)
Carbon fixation
PQQ-dependent dehydrogenase, methanol/ethanol family
Formaldehyde-activating enzyme
NAD(P)-dependent methylenetetrahydromethanopterin dehydrogenase MtdB
Methenyltetrahydromethanopterin cyclohydrolase Mch
Heterodisulfide reductase/formate dehydrogenase subunit gamma/DUF4399 domain-containing protein
Formate dehydrogenase
Formate dehydrogenase family accessory protein FdhD
Bifunctional protein FolD
NADH-quinone oxidoreductase subunit F
C1 oxidation
Transketolase
NADP-dependent phosphogluconate dehydrogenase/glucose-6-phosphate dehydrogenase
Pentose phosphate pathway
Fructose-bisphosphate aldolase
Triosephosphate isomerase
Pyruvate kinase
Glycolysis
Phosphoenolpyruvate synthase
Pyruvate, phosphate dikinase
Acyl-CoA dehydrogenase/oxaloacetate decarboxylase subunit alpha
Aldehyde dehydrogenase
PEP carboxykinase (GTP)/DUF302 domain-containing protein
PEP carboxylase
Dihydrolipoyl dehydrogenase
Bifunctional acyl-CoA synthetase/GNAT family N-acetyltransferase
Pyruvate metabolism
Aconitate hydratase A
Aconitate hydratase B
2-oxoglutarate dehydrogenase E1 component
Succinate-CoA ligase
Acetate-CoA ligase
Succinate dehydrogenase flavoprotein subunit
Fumarate hydratase
Malate dehydrogenase
Isocitrate lyase
Citric acid cycle
TRAP transporter large permease subunit
C4-dicarboxylate ABC transporter
C4-dicarboxylate ABC transporter substrate-binding protein
TRAP transport
Propionyl-CoA synthetase
Malyl-CoA lyase
Serine hydroxymethyltransferase 1
Hypothetical alpha amylase
Phosphomannomutase
Beta-(1-3)-glucosyl transferase
O
P
Y
AA
CC
V
W
X
Specimen
B
Carbon metabolism-related transcript clusters mapped to Spirochaeta-like species
Ribose-5-phosphate isomerase
Ribulose-phosphate 3-epimerase [Caldilinea aerophila]
Phosphoribosylpyrophosphate synthetase
Transketolase
Pentose phosphate pathway
6-phosphofructokinase
Carbohydrate kinase/fructokinase
Enolase
Glycolysis
Pyruvate dehydrogenase
Pyruvate dehydrogenase (acetyl-transferring), homodimeric type
Pyruvate, phosphate dikinase
Dihydrolipoyl dehydrogenase
Pyruvate metabolism
Citrate/2-methylcitrate synthase
Isocitrate dehydrogenase (NADP(+))
Succinate--CoA ligase subunit alpha
Succinate--CoA ligase subunit beta
Succinate dehydrogenase flavoprotein subunit
Citric acid cycle
C4-dicarboxylate ABC transporter permease [Deltaproteobacteria spp.]
TRAP transporter large permease subunit
TRAP dicarboxylate transporter, DctP subunit [Deltaproteobacteria spp.]
TRAP transport
4-alpha-glucanotransferase
Glucose-1-phosphate adenylyltransferase
Glycogen metabolism
Ribose ABC transport system, permease protein RbsC
D-xylose ABC transporter ATP-binding protein
Carbohydrate kinase
Ribose/xylose utilization
Transcriptional regulator, DeoR family
Thymidine phosphorylase
Deoxyribose-phosphate aldolase
Deoxyribose and deoxynucleoside catabolism
Branched-chain alpha-keto acid dehydrogenase subunit E2
2-oxo acid dehydrogenase subunit E2 [Firmicutes spp.]
Alpha-galactosidase
Formate C-acetyltransferase/glycerol dehydratase family glycyl radical enzyme/pyruvate formate lyase
Phosphate acetyltransferase
Sugar ABC transporter substrate-binding protein
Aldehyde dehydrogenase EutE
Acetolactate synthase, large subunit, biosynthetic type
UDP-glucose 4-epimerase GalE
LacI family transcriptional regulator
log2(TPM)
0
−5
−10
O
P
Y
AA
CC
V
W
X
Specimen

Figure 4.7. Log2-transformed TMM-normalized TPM of gene products of carbon metabolism-related transcript clusters mapped to the (A) thioautotrophic *S. floridana* symbiont (*Ca.* Thiodiazotropha endolucininae) and (B) *Spirochaeta*-like species. Transcript clusters with zero TPM values are represented as white cells.Abbreviations: PQQ, pyrroloquinoline-quinone; NADP, nicotinamide adenine dinucleotide phosphate; FolD, bifunctional methylene-H_4F dehydrogenase/methenyltetrahydrofolate cyclohydrolase; PEP, phosphoenolpyruvate; GTP, guanosine triphosphate; DUF, domain of unknown function; GNAT, GCN5-related N-acetyltransferase; TRAP, tripartite ATP-independent periplasmic transporter; ABC, ATP-binding cassette; Dct, dicarboxylate transport proteins; ATP; adenosine triphosphate; UDP, uridine diphosphate.

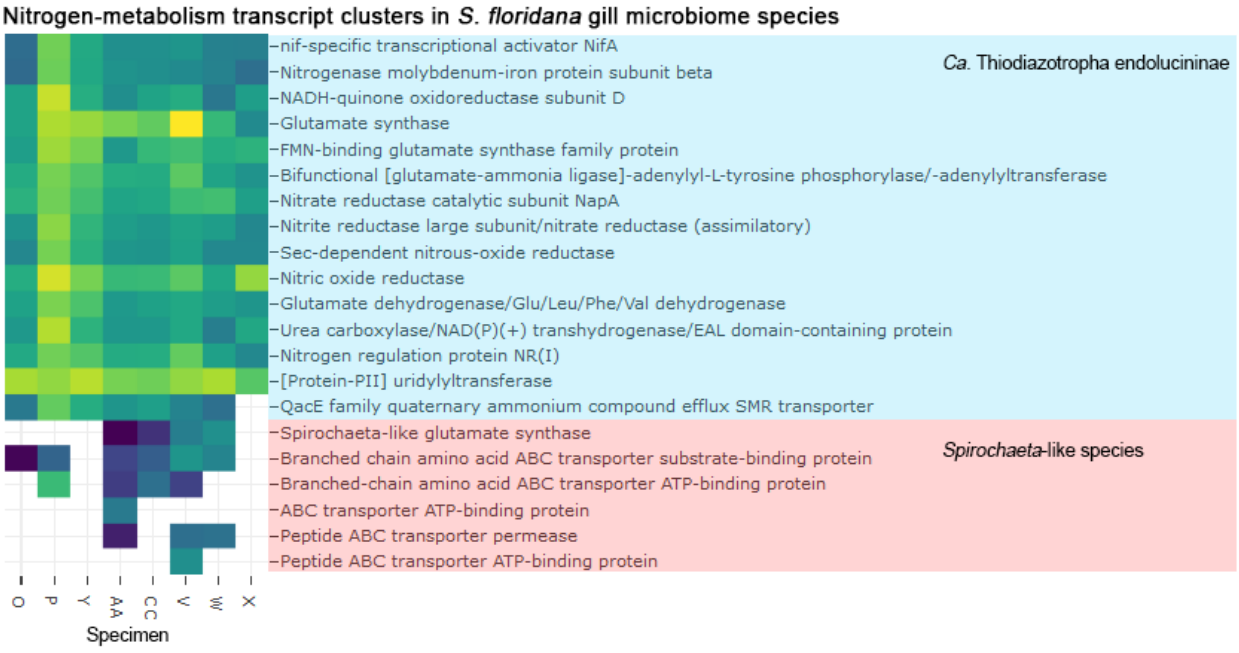


Figure 4.8. Log2-transformed TMM-normalized TPM of gene products of nitrogen metabolism-related transcript clusters mapped to the thioautotrophic *S. floridana* symbiont (*Ca.* Thiodiazotropha endolucininae) and *Spirochaeta*-like species. Transcript clusters with zero TPM values are represented as white cells. Abbreviations: *nif*, nitrogen fixation gene cluster; NADH, reduced nicotinamide adenine dinucleotide; FMN; flavin mononucleotide; Sec, secretory export protein; NADP, nicotinamide adenine dinucleotide phosphate; SMR, small multidrug resistance; ABC, ATP-binding cassette; ATP; adenosine triphosphate.

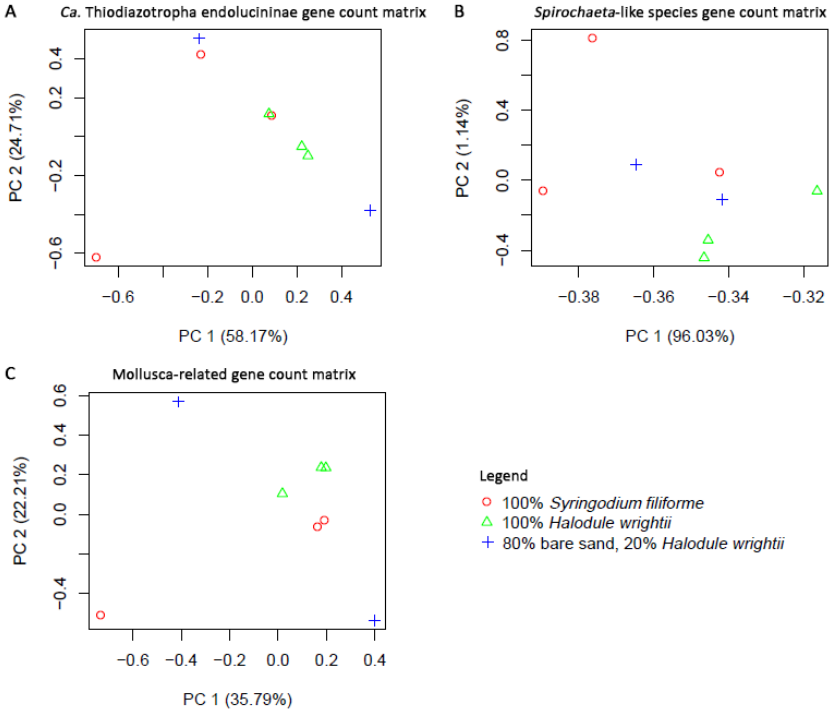


Figure 4.9. Principal component analyses of count matrices of transcript clusters mapped to genes from (A) *Ca.* Thiodiazotropha-like species, (B) *Spirochaeta*-like species, and (C) Mollusca species. The count matrices were processed to filter out genes with <10 mapped fragments, eliminate batch effects and normalized to $\log_2$ counts per million (CPM).

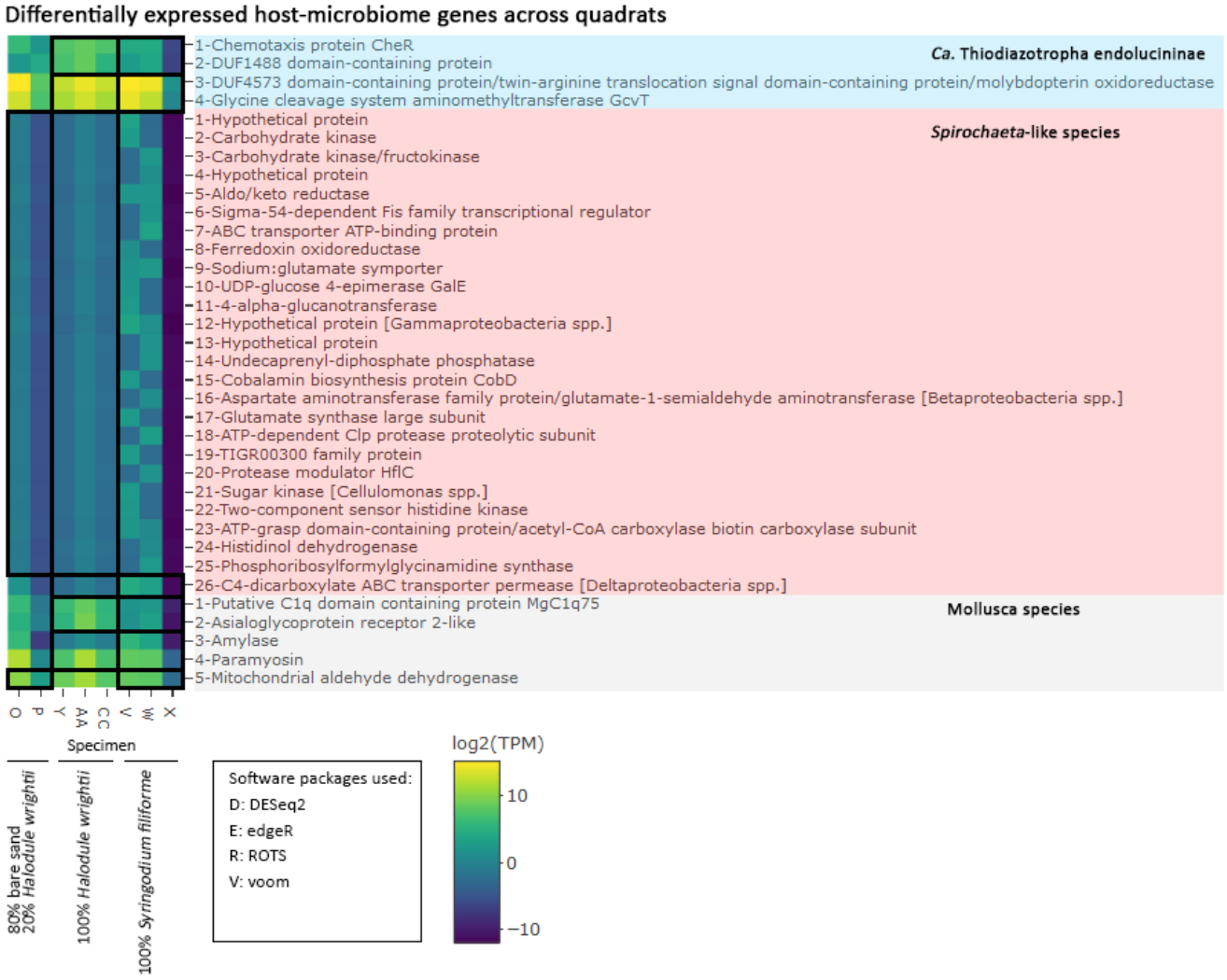


Figure 4.10. Differentially expressed ($p<0.05$, fold change≥ 2) genes mapped to *Ca.* Thiodiazotropha endolucininae, *Spirochaeta*-like species, and Mollusca species between *S. floridana* gill specimens collected from quadrats covered with 80% bare sand and 20% *Halodule wrightii* (T6/25m), 100% *Halodule wrightii* (T7/15m), and 100% Syringodium filiforme (T7/40m). Abbreviations: DUF, domain of unknown function; ABC, ATP-binding cassette; ATP; adenosine triphosphate; UDP, uridine diphosphate.

Spirochaeta-like species functions

Genes for the transport of sugar (average 1±1 TPM), dicarboxylate acids (average 0.3±0.4 TPM), branched chain amino acids (average 0.2±0.6 TPM), phosphate (average 0.3±0.8 TPM), sodium (average 0.2±0.2 TPM), and other unspecified substrates (average 0.5±1 TPM) were among the 35 most highly expressed protein-coding transcript clusters mapped to the *Spirochaeta*-like species MAG (Figure 4.5b). Among these, dicarboxylate acid- and phosphate transport (PhoU)-related transcripts and their corresponding sequences in the *Spirochaeta*-like species MAGs had homologs belonging to deltaproteobacterial species, which suggests that they were binned incorrectly or horizontally transferred (Figure 4.5b). The PhoU transcript cluster was also homologous to a similar protein in the *P. pectinatus*-associated spirochete (Chapter II). A growth-related transcript cluster encoding large subunit ribosomal proteins and three other transcript clusters encoding 50S ribosomal proteins were also expressed in the species at average 0.06±0.06 TPM and 0.03±0.06 TPM, respectively. Carbon metabolism-related genes expressed in this species were involved in the pentose phosphate pathway (average 0.1±0.3 TPM), glycolysis (average 0.03±0.04 TPM), pyruvate metabolism (average 0.2±0.9 TPM), citric acid cycle (average 0.09±0.2 TPM), organic acids transport (average 0.2±0.3 TPM), glycogen metabolism (average 0.6±0.2 TPM), ribose transport (average 0.2±0.03 TPM), xylose transport (average 0.2±0.2 TPM), and deoxyribose and deoxynucleoside catabolism (average 0.09±0.2 TPM; Figure 4.7b). Besides branched chain amino acids, the *Spirochaeta*-species possibly imports peptides (average 0.06±0.1 TPM) for nitrogen and assimilates ammonium using glutamate synthase (average 0.1±0.2 TPM; Figure 4.8). The

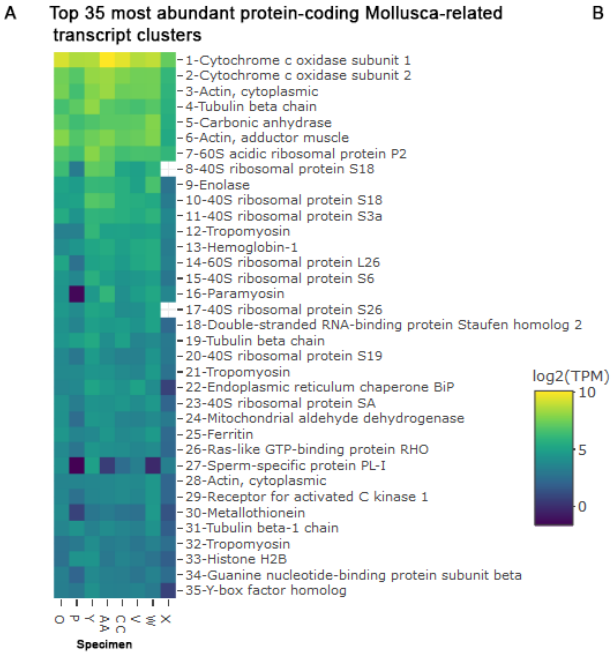
Spirochaeta-like species MAG contained genes for aerobic respiration with an unknown type cytochrome c oxidase and anaerobic respiration with an unknown electron acceptor, but these genes were not detected in the species' transcriptomes. The species encoded and/or expressed genes for the biosynthesis of proline (not in transcriptomes), alanine (average 0.09±0.2 TPM), leucine (average 0.2±0.3 TPM), threonine and homoserine (average 0.03±0.09 TPM), and aromatic compounds (not in transcriptomes). B vitamin biosynthesis potential in the *Spirochaeta*-like species included vitamins B1 (average 0.03±0.05 TPM), B6 (average 0.03±0.05 TPM), B7 (average 0.07±0.02 TPM), B9 (average 0.03±0.5 TPM), and B12 (average 0.03±0.09 TPM). The *Spirochaeta*-like species encoded components of the phosphate operon, among which the phosphate regulon sensor protein PhoR was expressed at average 0.02±0.07 TPM. Like transcriptomes of the *Ca.* Thiodiazotropha-like species, the *Spirochaeta*-like species transcriptomes' gene count matrix did not show apparent quadrat and vegetation-specific clustering patterns on the PCA plot (Figure 4.9b). A total of 128 genes, including 103 predicted by voom (Law *et al.*, 2014) and 25 predicted by both voom and ROTS (Suomi *et al.*, 2017), were differentially upregulated ($p<0.05$, ≥2 fold-change) in the 80% sand + 20% *Halodule wrightii*-covered quadrat compared to the 100% *H. wrightiii*-covered quadrat. The 25 commonly predicted DE genes were related to carbon, nitrogen, phosphate (two-component sensor histidine kinase mapped to phosphate regulon sensor protein PhoR), transport (sodium, glutamate and an unknown substrate), vitamin B12 (cobalamin) synthesis, histidine synthesis (histidinol dehydrogenase), transcriptional regulation, and protein degradation functions (Figure 4.10). Another incorrectly binned/horizontally-transferred transcript cluster

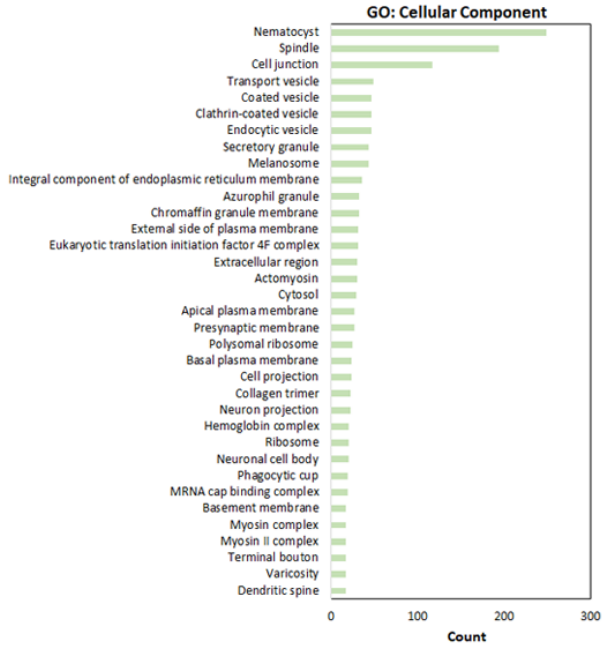
encoding a C4-dicarboxylate ABC transporter permease homologous to various Deltaproteobacteria species was differentially upregulated in the 100% *H. wrightiii*-covered quadrat compared to the 100% *S. filiforme*-covered quadrat (Figure 4.10).

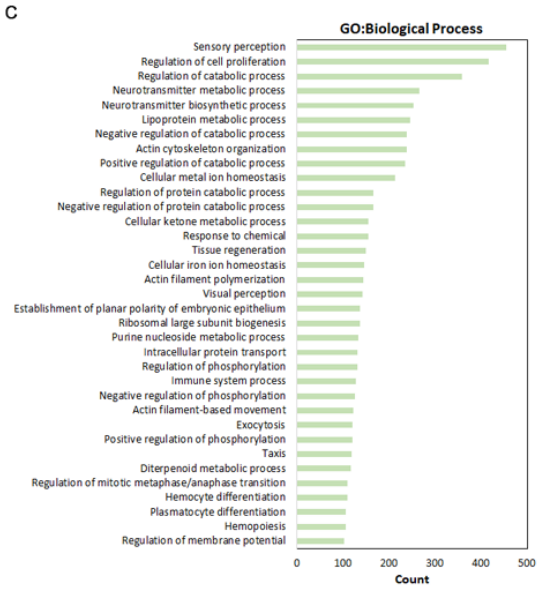
Host functions

The 35 most highly expressed protein-coding transcript clusters mapped to Swissprot (The UniProt Consortium, 2015)-annotated molluscan sequences were involved in aerobic respiration (cytochrome c oxidase; average 383±276 TPM), carbonic anhydrase (average 130±54 TPM), actin (average 98±97 TPM), tubulin (average 55±66 TPM), enolase (average 50±31 TPM), ribosomal protein (average 47±45 TPM), paramyosin (average 27±19 TPM), and tropomyosin (average 18±13 TPM) functions (Figure 4.11a). Accordingly, cellular component gene ontology (GO; Harris *et al.*, 2004) terms associated with actomyosin and myosin complexes, biological process GO terms associated with actin filament polymerization and movement, and molecular function GO terms related to microtubule motor activity and actin binding were among the 35 most frequently annotated from Mollusca-related transcript clusters (Figure 4.11b-d). A transcript cluster encoding hemoglobin 1 (average 32±15 TPM) and the cellular component GO term associated with the hemoglobin complex were the twelfth and 25th most abundant in the Mollusca-related transcriptomes, respectively (Figure 4.11a-b). In contrast, other transcript clusters encoding hemoglobin 3 (average 0.2±0.1 TPM) and globin (average 0.8±1 TPM) were expressed at >30x lower TPMs. The molecular function GO term for lysozyme activity was the 32nd most abundant in Mollusca-related transcript clusters (Figure 4.11d), and

genes encoding lysozymes 1 (average 0.8±0.5 TPM) and 3 (average 4±6 TPM) homologous to molluscan species were also identified in the gill metatranscriptomes. Other abundant Mollusca-related GO terms were associated with the nematocyst, spindle, cell junction and vesicular cellular compartments (Figure 4.11b), as well as sensory perception, regulation of cell proliferation, regulation of catabolic processes, and neurotransmitter biological processes (Figure 4.11c). The most frequently annotated Mollusca-related molecular function GO terms included binding functions for a variety of substrates such as phospholipids, neurotransmitters, retinoid, phosphatidylinositol-4,5-bisphosphate, peptide, nucleotide/nucleoside, histone, 11-cis retinal, signaling receptor, and hormone, along with channel activity, amylase activity, antioxidant activity, and hydrolase activity functions (Figure 4.11d). Like gene expression matrices of the *Ca.* Thiodiazotropha-like and *Spirochaeta*-like species, the transcript cluster count matrix of Mollusca-related genes did not show quadrat- and vegetation-specific grouping through PCA analysis (Figure 4.9c). Only five Mollusca-related DE genes ($p<0.05$, ≥ 2 fold-change) were predicted across quadrats with different vegetation coverages (Figure 4.10). Of these, transcript clusters encoding a putative C1q domain containing protein MgC1q75 and asialoglycoprotein receptor 2-like were upregulated, while an amylase-encoding gene was predicted by two algorithms to be downregulated in the 100% *H. wrightiii*-covered quadrat compared to the 100% *S. filiforme*-covered quadrat (Figure 4.10). Genes encoding paramyosin and mitochondrial aldehyde dehydrogenase were differentially upregulated in the 80% sand and 20% *H. wrightiii*-covered quadrat compared to the 100% *S. filiforme*-covered quadrat.


A
Top 35 most abundant protein-coding Mollusca-related transcript clusters
1-Cytochrome c oxidase subunit 1
2-Cytochrome c oxidase subunit 2
3-Actin, cytoplasmic
4-Tubulin beta chain
5-Carbonic anhydrase
6-Actin, adductor muscle
7-60S acidic ribosomal protein P2
8-40S ribosomal protein S18
9-Enolase
10-40S ribosomal protein S18
11-40S ribosomal protein S3a
12-Tropomyosin
13-Hemoglobin-1
14-60S ribosomal protein L26
15-40S ribosomal protein S6
16-Paramyosin
17-40S ribosomal protein S26
18-Double-stranded RNA-binding protein Staufen homolog 2
19-Tubulin beta chain
20-40S ribosomal protein S19
21-Tropomyosin
22-Endoplasmic reticulum chaperone BiP
23-40S ribosomal protein SA
24-Mitochondrial aldehyde dehydrogenase
25-Ferritin
26-Ras-like GTP-binding protein RHO
27-Sperm-specific protein PL-I
28-Actin, cytoplasmic
29-Receptor for activated C kinase 1
30-Metallothionein
31-Tubulin beta-1 chain
32-Tropomyosin
33-Histone H2B
34-Guanine nucleotide-binding protein subunit beta
35-Y-box factor homolog
log2(TPM)
10
5
0
O
P
Y
AA
C
V
W
X
Specimen


B
GO: Cellular Component
Nematocyst
Spindle
Cell junction
Transport vesicle
Coated vesicle
Clathrin-coated vesicle
Endocytic vesicle
Secretory granule
Melanosome
Integral component of endoplasmic reticulum membrane
Azurophil granule
Chromaffin granule membrane
External side of plasma membrane
Eukaryotic translation initiation factor 4F complex
Extracellular region
Actomyosin
Cytosol
Apical plasma membrane
Presynaptic membrane
Polysomal ribosome
Basal plasma membrane
Cell projection
Collagen trimer
Neuron projection
Hemoglobin complex
Ribosome
Neuronal cell body
Phagocytic cup
MRNA cap binding complex
Basement membrane
Myosin complex
Myosin II complex
Terminal bouton
Varicosity
Dendritic spine
0
100
200
300
Count


C
GO:Biological Process
Sensory perception
Regulation of cell proliferation
Regulation of catabolic process
Neurotransmitter metabolic process
Neurotransmitter biosynthetic process
Lipoprotein metabolic process
Negative regulation of catabolic process
Actin cytoskeleton organization
Positive regulation of catabolic process
Cellular metal ion homeostasis
Regulation of protein catabolic process
Negative regulation of protein catabolic process
Cellular ketone metabolic process
Response to chemical
Tissue regeneration
Cellular iron ion homeostasis
Actin filament polymerization
Visual perception
Establishment of planar polarity of embryonic epithelium
Ribosomal large subunit biogenesis
Purine nucleoside metabolic process
Intracellular protein transport
Regulation of phosphorylation
Immune system process
Negative regulation of phosphorylation
Actin filament-based movement
Exocytosis
Positive regulation of phosphorylation
Taxis
Diterpenoid metabolic process
Regulation of mitotic metaphase/anaphase transition
Hemocyte differentiation
Plasmatocyte differentiation
Hemopoiesis
Regulation of membrane potential
0
100
200
300
400
500
Count

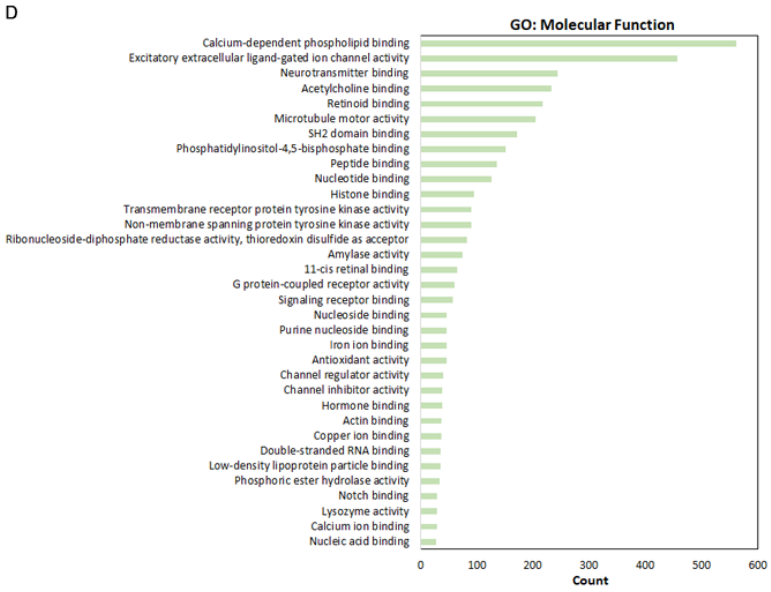

D
GO: Molecular Function
Calcium-dependent phospholipid binding
Excitatory extracellular ligand-gated ion channel activity
Neurotransmitter binding
Acetylcholine binding
Retinoid binding
Microtubule motor activity
SH2 domain binding
Phosphatidylinositol-4,5-bisphosphate binding
Peptide binding
Nucleotide binding
Histone binding
Transmembrane receptor protein tyrosine kinase activity
Non-membrane spanning protein tyrosine kinase activity
Ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
Amylase activity
11-cis retinal binding
G protein-coupled receptor activity
Signaling receptor binding
Nucleoside binding
Purine nucleoside binding
Iron ion binding
Antioxidant activity
Channel regulator activity
Channel inhibitor activity
Hormone binding
Actin binding
Copper ion binding
Double-stranded RNA binding
Low-density lipoprotein particle binding
Phosphoric ester hydrolase activity
Notch binding
Lysozyme activity
Calcium ion binding
Nucleic acid binding
0
100
200
300
400
500
600
Count

Figure 4.11. (A) Log2-transformed TMM-normalized TPM of gene products of the 35 most abundantly expressed protein-coding transcript clusters mapped to species from the phylum Mollusca and top 35 most represented gene ontology terms (Harris *et al.*, 2004) in the (B) cellular component, (C) biological processes and (D) molecular functions categories among mollusca-related transcript clusters. Transcript clusters with zero TPM values in (A) are represented as white cells.

Discussion

Despite experimental and sequencing-based studies on various aspects of the lucinid-bacteria symbiosis, little is known about the taxonomic and functional composition of lucinid gill microbiomes and their possible interactions with their hosts and their environments. In this study, we first compared similarities and differences of the *Stewartia floridana* gill microbiome with gill microbiomes of other lucinid species to evaluate 1) whether the thioautotrophic symbiont is homogenous in the gill community, and 2) whether other bacterial taxa are consistently present in the *S. floridana* gill microbiomes. We also assessed differences in host and symbiont gene expression across gill specimens sampled from three separate quadrats covered with 100% *H. wrightii*, 100% *Syringodium filiforme* and 80%-20% sand-*H. wrightii* mix. Like the *Codakia orbicularis* (genomes assembled from three individuals; n=3; *Ca.* Thiodiazotropha endolucinida; König *et al.*, 2016), *L. orbiculatus* (n=5; *Ca.* Thiodiazotropha endolucininae; Petersen *et al.*, 2016), and *P. pectinatus* (n=13; *Ca.* Sedimenticola endophacoides; Chapter II) gill microbiomes comprising monospecific clade A or clade C thioautotrophic symbionts, the *S. floridana*

gill microbiomes (n=6) harbored a homogenous thioautotrophic symbiont (*Ca.* Thiodiazotropha endolucininae) belonging to clade A, contrasting the higher level of thioautotrophic symbiont species and strain diversity observed in *Ctena orbiculata* (n=8; *Ca.* Thiodiazotropha endolucinida and *Ca.* Thiodiazotropha endolucinidaduo; Chapter III). Like gill microbiomes of *Euanodontia ovum* (Ball *et al.*, 2009), *Lucinoma aff. kazani* (Duperron *et al.*, 2007) and *P. pectinatus* (Chapter II), a *Spirochaeta*-like species that clustered with spirochetes in *P. pectinatus* and *Olavius* gutless marine worms (Blazejak *et al.*, 2005; Dubilier *et al.*, 1999) was identified in 72% of amplicon sequenced gill specimens (OTU2), all metagenomic gill libraries, and all metatranscriptomic gill libraries. Gene expression profiles of *S. floridana* and its *Ca.* Thiodiazotropha-like thioautotrophic symbiont were consistent across quadrats with different vegetation coverages, although a higher number of metabolic genes in the *Spirochaeta*-like species were differentially upregulated in the 80% sand + 20% *Halodule wrightii*-covered quadrat compared to quadrats entirely covered with *Halodule wrightii*.

The presence of monospecific thioautotrophic symbiont communities in *S. floridana*, as well as other lucinid species, including *Codakia orbicularis* (König *et al.*, 2016), *L. orbiculatus* (Petersen *et al.*, 2016), and *P. pectinatus* (Chapter II), suggests a strictly "one symbiont in one host" relationship in these species. This contrasts the "one host-multiple symbiont strains/species" interactions described in *Ctena orbiculata*, in which individuals can harbor more than one closely related thioautotrophic symbiont strain and/or species (Chapter III). Because lucinid bivalves acquire their symbionts environmentally, observed inter-host differences in thioautotrophic gill symbiont diversity

patterns may be related to symbiont recognition and acquisition mechanisms, lucinid species diversity in the habitat, or bacterial diversity of taxonomically and functionally similar gammaproteobacterial strains and species in lucinid habitats. These possibilities can be further tested with cross-infection experiments (Gros *et al.*, 2003b; Brissac *et al.*, 2009; Caro *et al.*, 2009), diversity analyses of environmental samples at the study site, and with additional sampling and comparisons of previously studied lucinid species from different sites and habitats.

With the increasing number of lucinid symbiont MAGs assembled to date, in this study, we were able to construct a robust phylogenomic tree based on 23 single-copy marker genes shared across these MAGs for higher species- and strain-level resolution. Our phylogenomic analysis of lucinid symbionts revealed four sequence-discrete phylogenomic clades, with each clade likely corresponding to a single species. The four proposed lucinid symbiont species, *Sedimenticola endophacoides* (Chapter II), *Ca.* Thiodiazotropha endolucinida (König *et al.*, 2016; Chapter III), *Ca.* Thiodiazotropha endolucinidaduo (Chapter III), and *Ca.* Thiodiazotropha endolucininae (Petersen *et al.*, 2016; this study) belong to two genera. So far, the genus *Ca.* Thiodiazotropha comprised clade A lucinid symbionts (classification scheme in Brissac *et al.*, 2011), while the genus *Sedimenticola* includes only the clade C *P. pectinatus* symbiont. Although clade A lucinid symbionts were previously thought to belong to a single species based on their 16S rRNA gene sequences (Brissac *et al.*, 2011), our metagenomic studies on thioautotrophic symbionts in *C. orbiculata* (Chapter III) and *S. floridana* (this study) uncovered a higher level of species diversity within clade A that could not have been resolved by the slow-

evolving 16S rRNA gene. Further, our phylogenomic analysis supports previous 16S rRNA gene-based findings that clade A lucinid symbiont species could be shared across different host taxa (Gros *et al.*, 2003b; Brissac *et al.*, 2011; Brissac *et al.*, 2016). This raises interesting questions on lucinid-microbiome-environment co-evolution, specifically, on the drivers of currently observed host-symbiont association patterns and the variability of these patterns across habitats. These can be investigated in the future with more extensive sequencing of lucinid symbionts across diverse and shared host taxa in different environments. These results also suggest that species naming schemes for lucinid symbionts, especially for taxonomically diverse clade A lucinid symbionts, have to take "one symbiont:multiple hosts" relationships into account. As such, we propose to replace the existing name of the *L. orbiculatus* and *S. floridana* symbiont species, *Ca.* Thiodiazotropha endoloripes (Petersen *et al.*, 2016), which inaccurately implies specific and exclusive association with *Loripes* spp., to the new name, *Ca.* Thiodiazotropha endolucininae, to reflect the general association of this symbiont species with lucinids from the Lucininae subfamily which includes but may or may not be limited to *L. orbiculatus* and *S. floridana*.

Like *L. orbiculatus*, *S. floridana* is commonly associated with seagrass beds and their chemosynthetic symbionts likely engage in thioautotrophic and diazotrophic functions that can benefit their surrounding seagrass habitats (Fisher and Hand, 1984; Meyer *et al.*, 2008; van der Heide *et al.*, 2012; Petersen *et al.*, 2016; Sanmartí *et al.*, 2018). Closely related thioautotrophic symbiont strains from both bivalve species were also genetically capable of hydrogen oxidation, mixotrophy, assimilatory and dissimilatory

denitrification, urea decomposition, and aerobic respiration. Unlike the *L. orbiculatus* symbiont, *Ca.* Thiodiazotropha endoloripes (Petersen *et al.*, 2016), however, the thioautotrophic *S. floridana* symbiont encoded and expressed genes from a conserved C1 compound oxidation gene cluster previously described in thioautotrophic gill symbionts from *Codakia orbicularis* and *Ctena orbiculata* (Chapter III). This metabolic difference could reflect symbiont adaptation or host selection for C1 compound oxidation functions that could be related to C1 compound concentrations in certain habitats. However, the concentrations of C1 compounds including methanol, formaldehyde, and formate were not measured in our study site. Hence, we currently do not know how the availability of C1 compounds in lucinid habitats affects the presence, absence and activity of C1 compound oxidation genes in lucinid symbionts. Low levels of methane were previously detected in porewaters from the *Ctena orbiculata* sampling site where C1-oxidizing thioautotrophic symbionts were sequenced and posited to be plant-derived (Chapter III). As with mutualisms of methylotrophic bacteria with strawberry plants and seagrasses (Abanda-Nkpwatt *et al.*, 2006; Crump *et al.*, 2018), C1-oxidizing lucinid endosymbionts could remove phytotoxic methanol from their surrounding seagrasses as part of their multi-faceted facilitative interactions with their hosts and their habitats (Chapter III). Controlled aquarium experiments will be useful in validating components of lucinid-microbiome-habitat interactions and establishing causal connections between host-symbiont functions and habitat modifications.

The low abundances of one or more closely-related *Spirochaeta*-like species consistently detected in the *S. floridana* gill microbiome corroborate previous reports of

spirochetes associated with marine organisms. In diverse marine ecosystems, spirochete species exist as free-living forms where they cycle sulfur and carbon compounds (Harwood and Canale-Parola, 1984; Breznak and Warnecke, 2008; Stephens *et al.*, 2008; Dong *et al.*, 2018) and host-associated forms where they inhabit tissues of red corals (van de Water *et al.*, 2016), the dorsal surface of the vent polychaete *Alvinella pompejana* (Campbell and Cary, 2001), the cuticle-epidermis space in *Olavius* gutless marine worms (Blazejak *et al.*, 2005; Dubilier *et al.*, 1999), crystalline styles in the digestive tracts of marine bivalves (Husmann *et al.*, 2010), and gills of lucinid bivalves (Duperron *et al.*, 2007; Chapters II and Chapter III). Specifically, in lucinid bivalves, spirochetes have been reported in the gills of *Lucinoma aff. kazani* (Duperron *et al.*, 2007), *Euanodontia ovum* (Espinosa *et al.*, 2013), *P. pectinatus* (Chapter II), as well as a foot specimen of *Ctena orbiculata* (Chapter III). Despite their widespread distribution, the roles of spirochete species in marine environments remain largely unknown. *Spirochaeta*-like OTU2 and OTU281-associated species in *S. floridana* were distantly related to the *Spirochaeta*-like species in the *P. pectinatus* gill microbiome and both formed sister groups to intracellular spirochete species in *Olavius* gutless marine worms (Blazejak *et al.*, 2005; Dubilier *et al.*, 1999). In comparison, spirochetes identified in *L. aff. kazani* (Duperron *et al.*, 2007) and the foot of *C. orbiculata* (Chapter III) did not belong to this clade and likely have different evolutionary origins. Currently, the phylogenetic positions of the *L. aff. kazani*-associated (Duperron *et al.*, 2007) and *C. orbiculata*-associated (Chapter III) spirochete species in relation to the *S. floridana*-associated *Spirochaeta*-like species could not be resolved using the V4 region of the 16S rRNA gene. *Spirochaeta*-like MAGs and transcriptomes in *S.*

floridana and *P. pectinatus* (Chapter II) contained genes that use sugars and other carbon sources not typically used by the thioautotrophic symbionts. The *S. floridana*-associated *Spirochaeta*-like MAG and transcriptomes also included genes for the transport of branched chain amino acids and peptides. The presence and expression of these genes suggest potential scavenging, nitrogen cycling, and carbon cycling roles of spirochete species in the gills of *S. floridana*. DE analyses of *Spirochaeta*-like transcript clusters revealed upregulation of carbon, nitrogen, phosphate, transport, and synthesis functions in the 80% sand + 20% *Halodule wrightii*-covered quadrat compared to the 100% *Halodule wrightiii*-covered quadrat, suggesting that these species could also be actively metabolizing substrates originating from the sediments. Because the prevalent *Spirochaeta*-like OTU 2 were not detected in every sequenced *S. floridana* gill specimen, we speculate that these species are facultative members of the gill microbiome. As with thioautotrophic lucinid symbionts, the spirochete species could be acquired by the clams from the environment. The latter could be sediment-dwelling bacteria trapped or enriched in the host gills that participate in commensal, amensal, parasitical or mutualistic relationships with the host, thioautotrophic symbionts and/or the surrounding habitat.

Despite the presence of spirochete-related sequences in gill microbiomes of *S. floridana*, in-depth analyses of the spirochetes' functions and differential gene expression were limited by the incompleteness of the *Spirochaeta*-like MAG, low numbers of metatranscriptomic reads mapping to the *Spirochaeta*-like MAG, and other general limitations pertaining to the rapid changes in gene expression before tissue fixation and quality of transcript annotations discussed at length in Chapter II. The same limitations

also apply to metagenomic and metatranscriptomic analyses of thioautotrophic symbiont functions in *S. floridana*. Additionally, as previously discussed in Chapter II, microscopic evidence is necessary to determine the location of these spirochete species in relation to the thioautotrophic symbionts in the gills for further inferences on species-species interactions. Microbiome analyses of sediment samples from the sampling sites, which were not performed in this study, would also enable meaningful comparisons of the relative abundances and phylogenetic diversities of the horizontally acquired thioautotrophic symbiont and spirochete species within and outside the gill environment. Additional sequencing and experimental efforts focusing on free-living and host-associated spirochete species in marine ecosystems would also greatly contribute to our understanding of their phylogeny, functions and ecological roles in general.

DE analyses of *Ca.* Thiodiazotropha-like and host-related genes from gill specimens collected from *Halodule wrightii*, *Syringodium filiforme* and sand-*Halodule wrightii* covered quadrats revealed low numbers of host and symbiont-related DE genes. This could be due to the lack of statistically significant differences in porewater geochemistry among quadrats within the sampling site and/or tight host regulation of the intracellular gill environment. Our results contrast previous findings where potential positive associations between lucinid abundances and seagrass were observed. At St Joseph Bay north of our sampling site, *S. floridana* population densities were not statistically different between *T. testudinum* and *R. maritama* seagrass beds; however, no live *S. floridana* specimens were found in entirely seagrass-free areas (Fisher and Hand, 1984). Another study on *Loripes orbiculatus* abundances in *Cymodocea nodosa* (seagrass)

meadows in the Mediterranean Alfacs bay reported statistically higher lucinid abundances in vegetated compared to bare sediments (Sanmartí *et al.*, 2018). In vegetated sediments, a negative correlation between *L. orbiculatus* abundances with sediment organic matter content was found and may be related to seagrass phenotypic variations (Sanmartí *et al.*, 2018). Specifically, *C. nodosa* in organic-rich sediments have less complex root systems than organic-poor sediments (Sanmartí *et al.*, 2018). On the other hand, DE analyses on *C. orbiculata* on a limited number of replicates showed upregulation of the thioautotrophic symbiont-related *dsrC* gene encoding a sulfurtransferase in an algae-covered quadrat compared to a seagrass-covered quadrat and upregulation of host-associated aerobic respiration, oxidative stress-stimulated apoptosis, tricarboxylic cycle/electron transport chain and mitochondrial sulfide oxidation functions in the seagrass-covered quadrat compared to the algae-covered quadrat possibly relevant to the three-way lucinid-symbiont-seagrass symbiosis (van der Heide *et al.*, 2012; Reynolds *et al.*, 2014; Chapter III). Nevertheless, there is a dearth of quantitative evidence correlating vegetation coverage, habitat geochemistry, lucinid abundances and host-symbiont gene expression, necessitating further system-level investigations integrating field measurements, sequencing data and laboratory experiments to validate the lucinid-symbiont-seagrass symbiosis model.

Besides DE host-related transcripts, commonly highly expressed host-related transcripts in *S. floridana* involved in aerobic respiration, cytoskeletal proteins, ribosomal proteins, lysozyme, carbonic anhydrase, and sulfide-reactive hemoglobin 1 functions (Kraus and Wittenberg, 1990) were in line with previous observations in *C. orbiculata*

(Chapter III) and *P. pectinatus* (Chapter II). Like *C. orbiculata* but unlike *P. pectinatus*, expression levels of oxygen-reactive hemoglobin 2 and 3 (Kraus and Wittenberg, 1990) were absent to low in *S. floridana* and may reflect high intracellular oxygen concentrations in the gills (Chapter III). Similarly, Mollusca-related GO terms associated with antioxidant activity were among the most abundantly annotated in the gills of *S. floridana*. GO terms associated with vesicular cellular compartments, as well as those associated with sensory perception and neurotransmitter functions, were among the most frequently annotated in *S. floridana*. Whether and how functions pertaining to endocytosis, vesicular transport (Cooper, 2000) and signaling are linked to symbiont acquisition, host-microbiome transport and host-microbiome communications at the molecular level remain to be elucidated with experimental approaches.

In summary, consistent with previous analyses on lucinid gill microbiome diversity (Brissac *et al.*, 2016; Chapters II and III), the observed taxonomic and functional diversity in the core *S. floridana* gill microbiome strengthen the concept of heterogeneous, rather than homogenous, lucinid gill microbiomes comprising communities of single or multiple thioautotrophic symbiont strains and species and/or non-thioautotrophic bacterial members (Chapters II and III). Through cross-microbiome comparisons, our results revealed strain- and species-level diversity in clade A lucinid symbionts, which expands current knowledge on intra-population, intra-host and inter-host variability of lucinid gill microbiome structures and functions. Although we did not identify high numbers of host- and thioautotrophic symbiont-related DE genes across quadrats with different vegetation coverages, we propose that the *Spirochaeta*-like species may participate in interactions (of

an unknown nature) with lucinid bivalves and surrounding bare sand sediments that probably involve transport, biosynthesis and cycling of a variety of nutrients including branched-chain amino acids and carbon, nitrogen, and phosphate compounds. Our study highlights the utility of ecosystem-based approaches in investigating the intricate interplay between lucinid bivalves, their symbionts and their surrounding habitats, while raising the need for further studies on lucinid gill microbiomes from diverse host species and habitats to better understand taxonomic and functional complexities in the lucinid-bacteria symbiosis.

CHAPTER FOUR

CONCLUSION

To summarize, this dissertation uses a combination of 16S rRNA, metagenomic, metatranscriptomic, PCR, and qPCR analyses to characterize the gill microbiomes of three coastal lucinid species, *Phacoides pectinatus*, *Ctena orbiculata*, and *Stewartia floridana*. To infer potential lucinid-microbiome interactions, gill microbiome gene expression was analyzed in relation to host gene expression. Additionally, to better understand spatial and micro-environment controls on host-microbiome gene expression, gill metatranscriptomic profiles of *C. orbiculata* and *S. floridana* were analyzed within the context of available environmental and/or geochemical data collected by our collaborators.

Despite intensive research on lucinid bivalves and their symbionts, the structure, functions, and interactions of lucinid gill microbiomes, which may contain permanent and/or transient bacterial members, are currently under-studied. Analyses of *P. pectinatus*, *C. orbiculata*, and *S. floridana* in this dissertation reveal unprecedented taxonomic and functional heterogeneity in their gill microbiomes that contrasts conventional assumptions of homogenous lucinid gill microbiomes with only monospecific chemosynthetic symbiont cultures. This dissertation also highlights inter-host species similarities and differences in gill microbiome structures and functions, which may reflect pathways essential for symbiosis, host-symbiont co-evolution processes, differences in symbiont acquisition and selection mechanisms, and/or variations in host intracellular/extracellular environments.

The gill microbiomes of all three lucinid species studied in this dissertation are dominated by high relative abundances of their chemosynthetic symbionts from symbiont gammaproteobacterial clades A and C. Thioautotrophic lucinid symbionts sequenced in this study belong to four sequence-discrete phylogenomic clades generally equivalent to species. These symbiont species were assigned the proposed names *Sedimenticola endophacoides* (*P. pectinatu*s thioautotrophic symbiont; Chapter II), *Ca.* Thiodiazotropha endolucinida (*Codakia orbicularis* and *Ctena orbiculata* thioautotrophic symbionts; König *et al.*, 2016; Chapter III), *Ca.* Thiodiazotropha endolucinidaduo (*Ctena orbiculata* thioautotrophic symbionts; Chapter III), and *Ca.* Thiodiazotropha endolucininae (to replace *Ca.* Thiodiazotropha endoloripes for *L. orbiculatus* and *S. floridana* thioautotrophic symbionts; Petersen *et al.*, 2016; Chapter IV). Results from this dissertation show the clade C *P. pectinatus* symbiont to belong to a genus (*Sedimenticola*) separate from other clade A lucinid symbionts (*Ca.* Thiodiazotropha). Additionally, the results highlight that species diversity of clade A lucinid symbionts, which now consists of three species rather than a single species, is higher than previously concluded using 16S rRNA gene sequences (Brissac *et al.*, 2011).

Comparative analyses of the genetic repertoire of *Ca.* Sedimenticola endophacoides, *Ca.* Thiodiazotropha endolucinida, *Ca.* Thiodiazotropha endolucinidaduo, and *Ca.* Thiodiazotropha endolucininae revealed commonalities in their core metabolic pathways that are in line with those discovered in gammaproteobaterial chemosynthetic marine symbionts. As with all other chemosynthetic marine symbionts reviewed in Kleiner *et al.* (2012), these thioautotrophic lucinid symbiont species potentially use *sox* (lacking

soxCD genes) and *dsr-apr-sat* pathway (Ghosh and Dam, 2009; Friedrich *et al.*, 2001) for the oxidation of sulfide and thiosulfate. Sulfide oxidation using Sqr enzymes (Marcia *et al.*, 2010; Eddie and Hanson, 2013; Shuman and Hanson, 2016) is also a common feature of thioautotrophic lucinid symbionts and some chemosynthetic marine symbionts, such as those in vesicomyid and *Bathymodiolus* spp. bivalves and those in *Riftia pachyptila* and *Tevnia jerichonana* deep-sea tubeworms (Kleiner *et al.*, 2012). Besides reduced sulfur compounds, hydrogen can also be used as a potential energy source in *Ca.* Sedimenticola endophacoides and *Ca.* Thiodiazotropha spp., along with some chemosynthetic marine symbionts in the bivalves *Bathymodiolus* spp. (Petersen *et al.*, 2011) and *Solemya velum* (Dmytrenko *et al.*, 2014), the gutless marine worm *Olavius algarvensis* (Woyke *et al.*, 2006; Kleiner *et al.*, 2015), the scaly-foot snail *Crysomallon squamiferum* (Nakagawa *et al.*, 2014), and the vent shrimp *Rimicaris exoculata* (Petersen *et al.*, 2011; Kleiner *et al.*, 2012). All chemosynthetic marine symbionts, including those from lucinid clams, encode autotrophy-related genes involved in the Calvin-Benson-Bassham cycle (Kleiner *et al.*, 2012). Unlike *R. pachyptila* symbionts that are also capable of autotrophy with the reductive tricarboxylic acid cycle (Markert *et al.*, 2007; Markert *et al.*, 2011; Gardebrecht *et al.*, 2012), the Calvin-Benson-Bassham cycle appears to be the sole autotrophic mechanism in chemosynthetic lucinid symbionts. Besides autotrophy, mixotrophy-related genes were detected in *Ca.* Sedimenticola endophacoides and *Ca.* Thiodiazotropha spp. (Petersen *et al.*, 2016), as well as chemosynthetic symbionts associated with marine organisms such as *O. algarvensis* (Woyke *et al.*, 2006), *R. pachyptila*/*T. jerichonana* (Kleiner *et al.*, 2012; Gardebrecht *et al.*, 2012), and *S. velum* (Dmytrenko *et al.*, 2014).

Candidatus Sedimenticola endophacoides and *Ca.* Thiodiazotropha spp. are potentially capable of co-respiring with oxygen and reduced nitrogeneous compounds, and nitrate respiration has similarly been documented in chemosynthetic marine symbionts associated with *O. algarvensis* and *R. pachyptila*/*T. jerichonana* (Kleiner *et al.*, 2012). Nevertheless, experimental approaches like those described in (Hentschel *et al.*, 1993; Hentschel and Felbeck, 1995; Hentschel *et al.*, 1996; Duplessis *et al.*, 2004a) are required to further ascertain the primary electron acceptor used by lucinid symbionts examined in this dissertation for respiration. Other common metabolic genes encoded and expressed by *Candidatus* Sedimenticola endophacoides and *Ca.* Thiodiazotropha spp. from this dissertation include those involved in phosphate and iron transport, biosynthesis of all twenty essential amino acids and vitamins B1, B2, B6, B7, B9, bacterial secretion systems, pilus and flagellar functions, and chemotaxis. Some of these functions may be relevant to host-microbiome-environment interactions and can be further investigated using protein-based, metabolite-based and experimental studies.

While the metabolic similarities of chemosynthetic lucinid symbionts provide useful insights on functions likely vital to their free-living and/or symbiotic lifestyle, inter-species metabolic differences uncovered in this dissertation may reflect variations in host, microbiome and/or environmental components that make up lucinid-microbiome-environment interactions. *Candidatus* Sedimenticola endophacoides belonging to clade C were taxonomically and functionally distinct from *Ca.* Thiodiazotropha spp. clade A lucinid symbionts. Members of the latter commonly encode and express at least the high-affinity type I Sqr (Eddie and Hanson, 2013; Shuman and Hanson, 2016), form IAq

RuBisCO that is more efficient at differentiating between oxygen and CO_2 (Tabita *et al.*, 2008), and the low-affinity aa3-type terminal oxidase (Pitcher and Watmough, 2004). In contrast, *Ca.* Sedimenticola enophacoides encoded and expressed the low affinity type VI Sqr (Eddie and Hanson, 2013; Shuman and Hanson, 2016), the less-discriminatory form II RuBisCO (Tabita *et al.*, 2008), and the high-affinity cbb3 type terminal (Pitcher and Watmough, 2004). These genomic differences suggest that, compared to *Ca.* Thiodiazotropha spp., *Ca.* Sedimenticola endophacoides appeared to be functionally adapted to sulfide-rich and oxygen-poor intracellular and/or extracellular environments. On the other hand, inter-species and inter-strain differences in the number of RuBisCO variants (form IAq and/or II) and terminal oxidases (cbb3-, aa3- and/or cytochrome d ubiquinol oxidases) encoded and expressed by *Ca.* Thiodiazotropha spp. suggest varying degrees of symbiont metabolic plasticity in response to intracellular/extracellular oxygen and CO_2 levels. C1-compound oxidation functions, detected in *Ca.* Thiodiazotropha endolucinida, *Ca.* Thiodiazotropha endolucinidaduo and *Ca.* Thiodiazotropha endolucininae in *S. floridana*, were not sequenced in *Ca.* Thiodiazotropha endolucininae in *L. orbiculatus* (Petersen *et al.*, 2016) and *Ca.* Sedimenticola endophacoides.

Despite common denitrification pathways in *Ca.* Thiodiazotropha spp. and *Ca.* Sedimenticola endophacoides, nitrogen assimilation mechanisms varied between symbiont genus, species and strains. The major difference between members of the *Ca.* Thiodiazotropha and *Sedimenticola* genera is the weak evidence supporting nitrogen fixation functions in the latter. Urea hydrolysis capability was predicted in *Ca.* Sedimenticola endophacoides and *Ca.* Thiodiazotropha endolucininae, but the

corresponding evidence is weak in *Ca.* Thiodiazotropha endolucinida and *Ca.* Thiodiazotropha endolucinidaduo. Also, assimilatory nitrate and nitrite reduction genes were annotated in *Ca.* Thiodiazotropha endolucininae, *Ca.* Thiodiazotropha endolucinidaduo and *Ca.* Thiodiazotropha endolucinida in *Ctena orbiculata*, but not *Ca.* Thiodiazotropha endolucinida in *Codakia orbicularis* (König *et al.*, 2016).

Inter-taxa metabolic differences among chemosynthetic symbionts observed in this dissertation may relate to lucinid-microbiome-environment co-evolution, which could be driven by geochemical conditions in lucinid habitats, lucinid-symbiont associations with their surrounding macro-vegetation, host metabolism (such as microbiome acquisition and selection mechanisms), and/or gill microbiome composition and functions. For example, the capability to fix nitrogen may be an evolutionary advantage for *Ca.* Thiodiazotropha spp. in seagrass habitats (König *et al.*, 2016; Petersen *et al.*, 2016), but not for *Sedimenticola endophacoides* in predominantly mangrove habitats. Similarly, C1 oxidation functions discovered in most *Ca.* Thiodiazotropha spp. may be dictated by the concentrations of C1 compounds, the relative abundances of free-living methanotrophs and methylotrophs and/or the need for methanol detoxification in their habitats. Technical variability arising from differences in metagenomic library preparation, assembly, binning and annotation methodologies across studies, which affect MAG completeness and quality, can also contribute to perceived inter-taxa, especially inter-strain, metabolic differences.

Chemosynthetic symbiont diversity within gill microbiomes of *P. pectinatus*, *C. orbiculata* and *S. floridana* also varied across host species. Unlike *P. pectinatus*, *S. floridana*, and other previously studied lucinid species with monospecific chemosynthetic

gill endosymbionts, the gills of *C. orbiculata* were colonized by at least two species, with two strains each, of chemosynthetic symbionts. These can possibly be explained by several hypotheses that can be tested in future, including 1) *C. orbiculata* acquires and selects for chemosynthetic symbionts with a lower stringency compared to other lucinid species, 2) turnover rate of *C. orbiculata* chemosymbionts is high because they are frequently replaced by closely related strains and species, and 3) the taxonomic diversity and relative abundances of free-living bacterial species closely related to lucinid chemosymbionts are higher in *C. orbiculata* habitats.

Besides the chemosynthetic symbionts, other taxa were also detected in the gill microbiomes of *P. pectinatus*, *Ctena orbiculata*, and *S. floridana.* Oceanospirillales species from different genera were detected consistently in the gill microbiomes, metagenomes, and metatranscriptomes of *P. pectinatus* (*Kistimonas*-like species) and gill microbiomes of *C. orbiculata* (*Endozoicomonas*-like species) at lower relative abundances compared to their chemosynthetic symbionts. *Spirochaeta*-like spp. also occurred in the gill microbiomes, metagenomes and metatranscriptomes of *P. pectinatus* and *S. floridana* at lower relative abundances compared to the Oceanospirillales species. The presence of these taxa supports previous reports of morphologically and/or taxonomically similar species in other lucinid gill microbiomes and suggests that lucinid gills may serve as permanent or temporary niches for Oceanospirillales and *Spirochaeta* species.

Beyond the taxonomic and functional diversity of lucinid gill microbiomes, this dissertation also sought to examine possible facilitative interactions between lucinid bivalves, their diverse gill microbiome members, and their micro-habitats. Exploratory DE

analyses performed on *C. orbiculata* revealed few differentially regulated symbiotic genes. The only symbiont-related DE gene of potential interest was *dsrC* encoding a sulfurtransferase that was up-regulated in an algae-covered quadrat. Meanwhile, *C. orbiculata*-related aerobic respiration, aerobic stress, electron transport, and mitochondrial sulfide detoxification functions were up-regulated in the seagrass-covered quadrat. In *S. floridana*, very few host and symbiont genes were differentially expressed between predominantly sand and predominantly seagrass-covered quadrats, while the *Spirochaeta*-like species showed upregulation of carbon, nitrogen, phosphate, transport, synthesis, transcriptional regulation, and protein degradation functions in predominantly sand-covered quadrats. The consistently small numbers of symbiont-related DE genes suggest that gill intracellular environments may be tightly regulated by lucinid bivalves as part of their homeostatic mechanisms. In comparison, the higher numbers of genes upregulated in the *Spirochaeta*-like species in predominantly unvegetated quadrats reveal potential associations of this species with bare sediments. On the other hand, respiratory and sulfide detoxification genes upregulated in *C. orbiculata* in the seagrass-covered quadrat offer preliminary evidence that these host functions are facilitated by the presence of seagrasses.

Although findings of this dissertation have provided useful insights on lucinid gill microbiome diversity and potential lucinid-microbiome-habitat interactions, various limitations challenge interpretation of the findings. First, functional inferences with metagenomic and metatranscriptomic data are heavily dependent on sequencing, assembly, binning, and annotation quality. Hence, incomplete MAGs and the low number of annotated transcripts assembled here may not present an accurate or complete picture of

the repertoire of genetic functions associated with the host and the microbiome species (discussed at length in Chapter II). In this dissertation, the presence and expression of few microbial genes of interest were further tested with PCR and qPCR. Even though it will be useful to perform PCR-based experiments on additional host and microbiome genes in the future to validate the results of metagenomic and metatranscriptomic analyses, not all mRNAs in a cell are necessarily being translated into proteins (Maquat *et al.*, 2010). Thus, protein and metabolite detection methods including Western blot, metaproteomics and metabolomics will be instrumental in confirming the activity of lucinid and microbiome pathways of interest, such as C1 oxidation in some chemosynthetic symbionts. Currently, annotation of host-related genes is also hampered by the lack of lucinid bivalve genomes. Future sequencing efforts should focus not only on lucinid-associated microbial genomes, but also on lucinid bivalve genomes. The availability of lucinid genomes would greatly facilitate research on the genetic and functional diversity of lucinid bivalves, as well as their co-evolution and interactions with their gill microbiome species and their surrounding habitats.

Second, the genetic and functional content of lucinid clam gills are heavily under-sampled to date. Currently, MAGs of clade A and C chemosynthetic symbiont species from only five host species have been sequenced (including those sequenced for this dissertation). These five species belong to three lucinid subfamilies, out of seven lucinid subfamilies (Taylor *et al.*, 2016) and >100 lucinid species (NCBI Resource Coordinators, 2016) identified to date. Furthermore, prior to this dissertation, no study has focused on comprehensive sequencing and characterization of the lucinid gill microbiomes. As such,

inferences made from cross-genome and cross-microbiome comparisons in this dissertation were centered on a small number of lucinid and symbiont taxa and may not be generally applicable to the entire Lucinidae family. For robust inter-host and inter-population comparisons of gill microbiome structures and functions, further sequencing of gill microbiomes of more lucinid species from various subfamilies and habitats is vital.

Third, although other non-chemosynthetic taxa were detected in the gill microbiomes of *P. pectinatus*, *Ctena orbiculata*, and *S. floridana*, without microscopic evidence, it is not possible to determine the localization of these gill microbiome species and the nature of their associations with the lucinid gill microbiomes. In this dissertation, microscopy was not performed on *C. orbiculata* and *S. floridana*. In *P. pectinatus*, FISH probes were successful in the detection of the chemosynthetic symbiont, but not in the detection of *Kistimonas*-like species and *Spirochaeta*-like species. This could be due to a variety of reasons pertaining to technical issues and inherent limitations of FISH itself. Another area of future research would be the optimization of FISH or more sensitive techniques, such as CARD-FISH (DeLong *et al.*, 1989), HCR (Dirks and Pierce, 2004) and/or electron microscopy for the visualization of other microbial taxa in lucinid gill microbiomes. Regardless of whether they are permanent or transient to the bivalves, results of this dissertation show that, as with microbiome species in various tissues of invertebrate and vertebrate hosts, lucinid gill microbiome species were transcriptionally active and may play nutrient cycling roles within the bivalves.

One of the objectives of characterizing the functions of lucinid bivalves and their microbiome species is to enable integrative analysis, in the context of environmental and/or

geochemistry data, of how lucinid-microbiome gene expression is affected by spatial, environmental and/or geochemistry parameters. Such ecological analyses can potentially improve lucinid, fisheries, seagrass, and mangrove conservation efforts (Johnson *et al.*, 2002; Meyer *et al.*, 2008; Reynolds *et al.*, 2014; Higgs *et al.*, 2016). In this dissertation, preliminary analyses of lucinid and microbiome gene expression in quadrats with varying vegetation coverages were conducted. However, statistical significance of the DE analyses was limited by the number of replicates in terms of the number of specimens collected and the environmental conditions tested. For field studies where environmental factors are unpredictable and uncontrollable, larger-scale sampling, preferably along an environmental gradient, in conjunction with -omics and microscopy studies, will be necessary to obtain sufficient replicates for robust statistical analysis and to better identify spatial/environmental/geochemical patterns that correlate with gene expression and other aspects of the lucinid-microbiome-environment relationships, such as host and microbiome associations, interactions, abundances, growth, fitness, morphology, and diversity. Alternatively, controlled aquarium experiments integrated with other -omics and microscopy approaches are useful in investigating not only environmental controls on host-microbiome physiology and diversity, but also host and microbiome effects on their surrounding environment. Symbiont cross-infection experiments, which have been successfully performed in some lucinid species (Brissac *et al.*, 2009; Caro *et al.*, 2009; Brissac *et al.*, 20160, can also help elucidate inter-host species differences in symbiont acquisition and selection mechanisms hypothesized in this dissertation.

Despite the limitations, this dissertation significantly advances existing knowledge on lucinid gill microbiome diversity, lucinid and microbiome functions and how these can potentially be influenced by their surrounding micro-environments. From a microbiology perspective, we now know that 1) lucinid gill microbiomes comprise of species- and function-diverse bacterial communities rather than a monospecific chemosymbiont culture, 2) Oceanospirillales and *Spirochaeta*-like species are facultatively associated with gills of multiple lucinid species, 3) clade A lucinid chemosymbionts belong to multiple species rather than a single species, 4) clade A and C lucinid chemosymbionts share common lithomixotrophic, biosynthesis, uptake and secretion systems functions, but differ in nitrogen assimilation, C1-compound oxidation, RuBisCO, sulfide oxidation, and aerobic respiration functions, 5) microbial genes differentially expressed in quadrats with varying vegetation coverages are related to sulfur oxidation (*Ca.* Thiodiazotropha endolucinidaduo OTU1) and nutrient cycling (*Spirochaeta*-like species in *S. floridana*). From a host perspective, this dissertation also uncovers a subset of the previously unsequenced lucinid transcriptome repertoire and how they can potentially be influenced by variability in vegetation coverages in their micro-environments.

Nevertheless, many gaps in knowledge on lucinid-microbiome-environment associations remain. Mainly, the intricate interplay and relative contributions of host metabolism, microbial functions, and environmental factors to lucinid-chemosymbiont associations, lucinid gill microbiome diversity and lucinid gill microbiome functions observed to date are still poorly understood. This dissertation proposes a variety of hypotheses relevant to this that should be validated with more extensive comparisons of

host, gill microbiome and environmental genetic, protein, and metabolite content across host taxa and habitats, combined with controlled aquarium and cross-inoculation experiments. Additionally, the nature of associations of non-chemosynthetic gill microbiome species with lucinid bivalves remain unknown and should be determined in future with high-resolution microscopy.

To conclude, findings of this dissertation revitalize the concept that lucinid gill microbiome communities are more functionally and taxonomically complex than previously thought and bring us a step closer towards understanding the many host-microbe-environment interactions possible within this remarkably multi-faceted symbiotic system.

APPENDICES

Appendix A: Computer commands and scripts used for data analysis

A1. Read trimming and sequence processing

1. Bash script to automate read trimming by cutadapt (removal of Illumina adaptors and quality trimming) and sickle (second round of quality trimming), followed by conversion of fastq files to fasta files for assembly (optional).

   The script assumes all input files to be in the working directory, where forward reads contain the label "R1" and ends with the extension.fastq.

```
for i in `ls *R1*fastq | awk -F "-" '{print $2}' | sed
"s/_.*$//g"`
do
cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTC -A
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -q 30 -o 30."$i".R1.fq -p
30."$i".R2.fq *"$i"*R1*.fastq *"$i"*R2*
.fastq
sickle pe -q 30 -f 30."$i".R1.fq -r 30."$i".R2.fq -t sanger
-o "$i".R1.fq -p "$i".R2.fq -s "$i".singles.fq
seqtk seq -A "$i"*L001_R1.fq  >"$i".R1.fa
seqtk seq -A "$i"*L001_R2.fq  >"$i".R2.fa
done
```

2. Bash one-liner to remove extra line breaks for sequences in fasta file. Output file will have exactly one line of header and one line of sequence.

```
cat input.fasta | sed "s/>.*$/#&#/g" | tr -d "\n" | tr "#"
"\n" | grep -v "^$" >output.fasta
```

A2. 16S rRNA gene analysis pipelines

1. 16S rRNA gene analysis pipeline using Mothur. Comments are in bold.

 # Combine forward and reverse reads into a single fasta file
> make.contigs(file=clam.file,processors=12)

 # Trim off forward and reverse primers, if necessary
> pcr.seqs(fasta=current,oligos=primer.oligos)

 # Contents of "primer.oligos" file
forward GCCGCGGTAA
reverse GGGTNTCTAAT

 # Trim sequences at Q=25 threshold
> trim.seqs(fasta=current,qfile=current,qaverage=25)
> remove.seqs(group=current,accnos=ctena.trim.accnos)
> screen.seqs(fasta=current,group=current,
summary=current,maxambig=0,maxlength=275)
> summary.seqs(fasta=current)
> unique.seqs(fasta=current)
> count.seqs(name=current,group=current)
> summary.seqs(count=current)

 # Download Silva v132 reference files
wget https://mothur.org/w/images/3/32/Silva.nr_v132.tgz
tar -zxvf Silva.nr_v132.tgz

 # Extract V4 region from Silva v132 reference fasta file
> pcr.seqs(fasta=silva.nr_v132.align,start=11894,end=25319)

 # Align sequences with reference sequences from Silva v132
> align.seqs(fasta=current,reference=../silva132/
silva.nr_v132.pcr.align,flip=T)
> summary.seqs(fasta=current,count=current)
> screen.seqs(fasta=current,count=current,summary=current,
start=13862,end=23444,maxhomop=8)
> summary.seqs(fasta=current,count=current)
> remove.seqs(accnos=current,count=current)
> summary.seqs(fasta=current,count=current)
> filter.seqs(fasta=current,vertical=T,trump=.)
> unique.seqs(fasta=current,count=current)
> pre.cluster(fasta=current,count=current,diffs=2)

 # Check for and remove chimeric sequences
> chimera.vsearch(fasta=current,count=current,dereplicate=t)
> remove.seqs(fasta=current,accnos=current)
> summary.seqs(fasta=current,count=current)

 # Assign taxonomy to each sequence using Silva v132 reference taxonomy
> classify.seqs(fasta=current,count=current,reference=
../silva132/silva.nr_v132.pcr.align,taxonomy=
../silva132/silva.nr_v132.tax,cutoff=0)

 # Remove non-bacterial and unknown sequences
> remove.lineage(fasta=current,count=current,
taxonomy=current,taxon=Chloroplast-Mitochondria-unknown-
Archaea-Eukaryota)
> summary.tax(taxonomy=current,count=current)

 # Compute distances between sequences
> dist.seqs(fasta=current,cutoff=0.03)

 # Cluster sequences *de novo* into OTUs
> cluster(column=current,count=current)

 # Make shared file using 0.01 (99% sequence identity) cutoff
> make.shared(list=current,count=current,label=0.01)

 # Count number of sequences in each sample (group)
> count.groups(shared=current)

 # Subsample each group to specific number of sequences
> sub.sample(shared=current,size=<**user-specified size**)

 # Compute alpha diversity measures for subsampled OTU table
> summary.single(shared=current)

 # Get OTU relative abundances in each subsampled group
> get.relabund(shared=current)

 # Export OTU table and taxonomy into BIOM format
> make.biom(shared=current,constaxonomy=current)

 # Get representative sequences from each OTU
> get.oturep(column=current,count=current,fasta=current)

2. R Alpha and beta diversity analysis pipeline using the PhyloSeq and other R packages. Comments are in bold.

 # Import phyloseq and related R libraries
library("phyloseq")
library ("ggplot2")

 # Import biom file created by Mothur (previous section)
clam=import_biom("clam.0.01.subsample.0.01.biom")

 # Import metadata
map <- import_qiime_sample_data("clam.design.txt")

 # Sample contents of "clam.design.txt" metadata file

SampleID	Type	Quadrat
1	Gill	1-2
2	Gill	1-2
3	Gill	1-2
4	Gill	1-2

 # Merge OTU tables, taxonomy and metadata
ste <- merge_phyloseq(clam,map)

 # Sample commands to plot alpha diversity
options(device=pdf)
plot_richness(ste,x="Type",shape="Type",color="Quadrat",measures=c("Chao1", "Shannon"))
dev.off()

 # Sample commands to plot beta diversity
options(device=pdf)
ord <- ordinate (ste,"PCoA","bray")
p = plot_ordination(ste,ord,type="samples",color="Tissue",shape="Type")
p
dev.off()

 # Change axis and plot titles
p = p + geom_point(size=4) + xlab("PC 1 (51.1%)") + ylab("PC2 (19.8%)") + ggtitle ("adonis F=208.79 (p=0.001***)")

 # Import R libraries for statistical analyses
library("vegan")
library("car")

 # Perform shapiro and adonis tests on a Bray-Curtis distance matrix using a specified metadata category (e.g. “Quadrat”)

```
df=as(sample_data(clam),"data.frame")
d=distance(clam,"bray")
shapiro.test(d)
adonis(d~Quadrat,df)
```

 # Example commands comparing statistical differences between alpha diversity measures (e.g. Shannon) – testing for normality

```
alpha<-read.table("alpha.txt",header=T, sep="\t")
shapiro.test(alpha$Shannon)
```

 # Example commands comparing statistical differences between alpha diversity measures (e.g. Shannon) for normally distributed data

```
var.test(alpha$Shannon~alpha$Type)
bartlett.test(alpha$Shannon~alpha$Type)
t.test(alpha$Shannon~alpha$Type)
fit<-aov(alpha$Shannon~alpha$Type)
summary(fit)
```

 # Example commands comparing statistical differences between alpha diversity measures (e.g. Chao1) for non-normally distributed data

```
leveneTest(alpha$Chao1,alpha$Type)
fligner.test(alpha$Chao1~alpha$Type)
wilcox.test(alpha$Observed~alpha$OTU)
pairwise.wilcox.test(alpha$Shannon,alpha$OTU,p.adjust.method="bonferroni",pa
ired=FALSE)
```

A3. Metagenomic binning

1. Bash Portable Batch Scheduling system (PBS) script to automate read mapping to assembled metagenome, conversion of sam output file to bam file, and binning with MetaBAT1 and MetaBAT2 on Clemson University's Palmetto cluster.

```
#!/bin/bash
#PBS -N metabat
#PBS -l
select=1:ncpus=16:mpiprocs=16:mem=120gb:interconnect=1g,wall
time=72:00:00
#PBS -j oe

module load samtools
module load boost
module load python/2.7.6

DIR="/scratch1/jslim"

/home/jslim/bowtie2-2.3.3.1/bowtie2-build
$DIR/LUC13015.AG.fasta $DIR/LUC13015AG
/home/jslim/bowtie2-2.3.3.1/bowtie2 -p 16 --very-sensitive-
local --dovetail -x $DIR/LUC13015AG -1 $DIR/LUC13015Y.R1.fq
-2 $DIR/LUC13015Y.R2.fq -S $DIR/LUC13015.sa
m -p 12

samtools view -bS $DIR/LUC13015.sam -o $DIR/LUC13015.bam -@
12
samtools sort $DIR/LUC13015.bam -o $DIR/LUC13015.sorted.bam
-@ 12
samtools index $DIR/LUC13015.sorted.bam -@12

/home/jslim/metabat1/jgi_summarize_bam_contig_depths --
outputDepth $DIR/LUC13015AG.depth.txt
$DIR/LUC13015.sorted.bam
/home/jslim/metabat2/metabat2 -i $DIR/LUC13015.AG.fasta -a
$DIR/LUC13015AG.depth.txt -o $DIR/LUC13015AGmetabat2 -t 12 -
m 1500
/home/jslim/metabat1/metabat -i $DIR/LUC13015.AG.fasta -a
$DIR/LUC13015AG.depth.txt -o $DIR/LUC13015AGmetabat --
sensitive -v --saveTNF $DIR/LUC13015.tnf --saveDistance
$DIR/LUC13015.dist -B 20 -t 12 -m 1500
```

A4. Pairwise average nucleotide identity (ANI) calculations

1. Perl script to automate pairwise ANI calculations in a set of input files with ".fa" extension in a folder (assigned to the variable $q and parsed into an array named @query). Output files generated will consist of the file names being compared and end with the extension .ani (e.g. file1.file2.ani).

```
#!/usr/bin/perl

$q=`ls *fa`;
@query=split(/\n/,$q);

$size=@query;

foreach my $n (@query) {
for ($index=0; $index<$size; $index++) {

system("/home/shared/ANIcalculator_v1/ANIcalculator -
genome1fna ".$n." -genome2fna ".$query[$index]." -outfile
".$n.$query[$index].".ani");

}
}
```

Example output file (e.g. file1.file2.ani)

```
GENOME1 GENOME2 ANI(1->2)       ANI(2->1)       AF(1->2)
AF(2->1)
spiro.1.fa      Stebin4.fa      69.36   69.58   0.13    0.24
```

2. Bash one-liner to concatenate .ani files generated by Perl script and calculate the average ANIs for each pairwise comparison.

```
cat *.ani | grep -v "GENOME1" | awk -F "\t" '{a=($3+$4)/2;
print $1,$2,a}' >ani.average
```

Example output file (e.g. ani.average)

```
19G.fa 22B.fa 74.705
19G.fa 22G.fa 74.85
19G.fa 4A.1.fa 75.24
19G.fa 4D.fa 74.73
19G.fa Codakia.symbiont.fasta 74.365
```

A5. Heatmap plotting in RStudio

1. R script to automate heatmap plotting with the heatmaply package, given an input matrix in tab-separated text file format within the specified working directory.

```
library('heatmaply')

setwd("G:/My Drive/clams/ANI")


ANI<-
read.table('ctena.AAI.txt',sep="\t",header=TRUE,row.names=1,
check.names=FALSE)
matrix<-data.matrix(ANI)
heatmaply(matrix, column_text_angle=90,key.title="Two-way
AAI") %>% layout(margin = list(l = 220, b = 220))
```

```
# Example input file (e.g. ctena.AAI.txt)

      P4    P5    P2    P1    P3
P4    100   99.99 67.87 99.99 0
P5    99.99 100   67.79 99.99 0
P2    67.87 67.79 100   67.84 0
P1    99.99 99.99 67.84 100   65.61
P3    0     0     0     65.61 100
```

A6. Metatranscriptomic analysis

1. Metatranscriptomic analysis pipeline using Trinity. Comments are in bold.

 # Transcript assembly: Run jellyfish, inchworm, and chrysalis steps
$TRINITY_HOME/Trinity --seqType fq --samples_file ctena.samples.txt --CPU 24 --max_memory 400G --output duo_trinity --no_run_inchworm

$TRINITY_HOME/Trinity --seqType fq --samples_file ctena.samples.txt --CPU 24 --max_memory 400G --output duo_trinity --no_run_chrysalis

$TRINITY_HOME/Trinity --seqType fq --samples_file ctena.samples.txt --CPU 24 --max_memory 400G --output duo_trinity --no_distributed_trinity_exec

$TRINITY_HOME/Trinity --seqType fq --samples_file ctena.samples.txt --CPU 24 --max_memory 400G --output duo_trinity

 # Abundance estimation with RSEM and Bowtie2
$TRINITY_HOME/util/align_and_estimate_abundance.pl --transcripts Trinity.fasta --est_method RSEM --aln_method bowtie2 --trinity_mode --prep_reference

 # Convert abundance estimates to TPM and TMM-normalized TPM distance matrices
$TRINITY_HOME/util/abundance_estimates_to_matrix.pl --est_method RSEM --gene_trans_map Trinity.fasta.gene_trans_map --out_prefix <choose_a_prefix> --name_sample_by_basedir <sample1_name>/RSEM.isoforms.results <sample2_name>/RSEM.isoforms.results

 # Bash one-liner to calculate average TPM across samples from the isoform/gene count matrix and sort the resulting count matrix by the average TPM values in descending order. In this case, the example input has 10 columns, where columns 2 ($2) to 11 ($11) contain the TPM values.
cat ctena.gene.TMM.EXPR.matrix | awk -F "\t" '{a=($2+$3+$4+$5+$6+$7+$8+$9+$10+$11)/10; print $1,$2,$3,$4,$5,$6,$7,$8,$9,$10,$11,a}' | sort -n -r -k 12,12 >ctena.sorted.gene.matrix

 # Remove batch effects from user-specified count matrix
$TRINITY_HOME/Analysis/DifferentialExpression/remove_batch_effects_from_count_matrix.pl --matrix ctena.gene.counts.matrix --batches_file batch.txt

 # Example batch file (e.g. batch.txt) for batch effects removal command
cond_1 4A
cond_1 4B

 # Compare replicates across samples and generate a sample correlation heatmap. The command takes in any count matrix (unnormalized, normalized, batch removed etc).
$TRINITY_HOME/Analysis/DifferentialExpression/PtR --matrix ctena.isoform.counts.matrix --min_rowSums 10 -s ctena.samples.txt --log2 --CPM --sample_cor_matrix

 # Compare replicates across samples and generate a 2-dimensional PCoA plot (--prin_comp 2):
$TRINITY_HOME/Analysis/DifferentialExpression/PtR --matrix ctena.isoform.counts.matrix --min_rowSums 10 -s ctena.samples.txt --log2 --CPM --center_rows --prin_comp 2

 # Bash one-liner to extract sequence information from RAST-annotated MAGs and write out as FASTA file
cat *.txt | awk -F "\t" '{print $2,$12}' | sed "s/^/>/g" | tr " " "\n" >ctena.pangenome.fasta

 # Deduplicate pangenome fasta file at 100% identity threshold with cd-hit
cd-hit -i ctena.pangenome.fasta -c 1 -o ctena.pangenome.dedup.fasta -T 0 -M 100000

 # Format Trinity-assembled transcripts into a blastn searchable database
makeblastdb -in ctena_trinity_out/Trinity.fasta -dbtype nucl -out trinity

 # Perform blastn searches against Trinity-assembled transcripts using RAST-annotated gene sequences as query and generate output in tabulated format.
blastn -query ctena.pangenome.dedup.fasta -db trinity -max_target_seqs 1 -num_threads 56 -outfmt 7 -out ctena.pangenome.blastn

 # Bash script to perform a basic keyword search in a user-provided genome annotation file, print out peg IDs containing the input keyword and their best-matching Trinity transcripts and their TMM-normalized TPM values. Intermediate files with peg IDs (e.g. keyword.peg) and Trinity transcript IDs (e.g. keyword.trinity) will be generated and have to removed periodically.

grep -i $1 pangenome.annotations | awk -F "\t" '{print $2}' | sed "s/$/[[:blank:]]/g" >$1.peg
grep -f $1.peg pangenome.onehit | awk -F "\t" '{print $1,$2}'
grep -f $1.peg pangenome.onehit | awk -F "\t" '{print $2}' | sed "s/_i.*$/[[:blank:]]/" >$1.trinity
grep -f $1.trinity pha.gene.TMM.EXPR.matrix

 # Output of bash script above

The output of the script will be printed on the screen (stdout) and not written to any file. The output can be easily redirected to a file using the ">" option on the command line when the script is executed.

First half of the output shows the peg ID and matching Trinity IDs, while second half of the output shows the Trinity IDs and the TPM values in each sample (in this case, there were 3 samples)

```
fig|6666666.229992.peg.1643 TRINITY_DN372328_c3_g2_i1
fig|6666666.270266.peg.151 TRINITY_DN84809_c0_g1_i1
fig|6666666.270266.peg.159 TRINITY_DN736615_c0_g1_i1
TRINITY_DN203370_c0_g1  0.000   0.000   0.201
TRINITY_DN315550_c0_g2  0.010   0.020   0.131
TRINITY_DN366491_c0_g2  0.472   0.101   1.096
```

 # Sample user-customizable GLOBALS section of configuration file for automating Trinotate's transcript annotation process

```
###############################################################################
# Globals. Specify resource locations and other templated parameter values
# Use format {__token__} when using token values in command strings.
# Other templated parameters are defined by the parent script.
###############################################################################

[GLOBALS]

#  ** edit the progs and dbs section to point to your local resources.

# progs
TRANSDECODER_DIR=$TRINOTATE_HOME/TransDecoder-TransDecoder-v5.1.0
BLASTX_PROG=blastx
BLASTP_PROG=blastp
SIGNALP_PROG=$TRINOTATE_HOME/signalp-4.1/signalp
TMHMM_PROG=$TRINOTATE_HOME/tmhmm-2.0c/bin/tmhmm
RNAMMER_TRANS_PROG=$TRINOTATE_HOME/util/rnammer_support/RnammerTranscriptome.pl
RNAMMER=$TRINOTATE_HOME/rnammer/rnammer
HMMSCAN_PROG=hmmscan

# dbs
SWISSPROT_PEP=$TRINOTATE_HOME/admin/uniprot_sprot.pep
PFAM_DB=$TRINOTATE_HOME/admin/Pfam-A.hmm
```

 # Create a SQlite database named "Trinotate" to hold all Trinotate annotation data
$TRINOTATE_HOME/admin/Build_Trinotate_Boilerplate_SQLite_db.pl Trinotate

 # Run Trinotate
$TRINOTATE_HOME/auto/autoTrinotate.pl --Trinotate_sqlite Trinotate.sqlite --transcripts Trinity.fasta --gene_to_trans_map Trinity.fasta.gene_trans_map --conf $TRINOTATE_HOME/auto/conf.txt --CPU 56

 # Run signalP for the prediction of signal peptides manually if Trinity assembly has >10,000 sequences. Split Trinity assembly into multiple files (e.g. 20,000 lines per file).
split -d -l 20000 Trinity.fasta.transdecoder.pep tsplit
mkdir Trinity.split
mv tsplit* Trinity.split
cd Trinity.split

 # Bash script to run signalP on each split file
for i in `ls tsplit*`
do
$TRINOTATE_HOME/signalp-4.1/signalp -f short -n $i.out $i
sleep 5
done

 # Merge all signalP output files into one file:
cat tsplit*.out >signalp.out
cd ..

 # Populate the "Trinotate" SQlite database with transcript sequences, protein sequences and gene/transcript relationships
$TRINOTATE_HOME/Trinotate Trinotate.sqlite init --gene_trans_map Trinity.fasta.gene_trans_map --transcript_fasta Trinity.fasta --transdecoder_pep Trinity.fasta.transdecoder.pep

 # Store blastp results in "Trinotate" SQlite database
$TRINOTATE_HOME/Trinotate Trinotate.sqlite LOAD_swissprot_blastp blastp.outfmt6

 # Store blastx results in "Trinotate" SQlite database
$TRINOTATE_HOME/Trinotate Trinotate.sqlite LOAD_swissprot_blastx blastx.outfmt6

 # Store Pfam domain entries in "Trinotate" SQlite database

$TRINOTATE_HOME/Trinotate Trinotate.sqlite LOAD_pfam TrinotatePFAM.out

 # Store transmembrane domain predictions in "Trinotate" SQlite database
$TRINOTATE_HOME/Trinotate Trinotate.sqlite LOAD_tmhmm tmhmm.out

 # Store signal peptide predictions in "Trinotate" SQlite database
$TRINOTATE_HOME/Trinotate Trinotate.sqlite LOAD_signalp Trinity.split/signalp.out

 # Output Trinotate annotation report. The –incl_pep and –incl_trans options add protein and transcript sequence data to the report. Output file will be an Excel sheet (e.g. trinotate_annotation_report.xls).
$TRINOTATE_HOME/Trinotate Trinotate.sqlite report --incl_pep --incl_trans >trinotate_annotation_report.xls

 # Format of Trinotate annotation report file. The filet has the following column headers:
0 #gene_id
1 transcript_id
2 sprot_Top_BLASTX_hit
3 RNAMMER
4 prot_id
5 prot_coords
6 sprot_Top_BLASTP_hit
7 custom_pombe_pep_BLASTX
8 custom_pombe_pep_BLASTP
9 Pfam
10 SignalP
11 TmHMM
12 eggnog
13 Kegg
14 gene_ontology_blast
15 gene_ontology_pfam
16 transcript
17 peptide

 # Example command to run differential expression analysis (e.g. using edgeR)
$TRINITY_HOME/Analysis/DifferentialExpression/run_DE_analysis.pl --matrix ctena.gene.counts.matrix.batch_eff_removal.matrix --method edgeR --samples_file metadata.txt --output output_dir

 # Run GO enrichment analysis – extract GO terms from Trinotate report
$TRINOTATE_HOME/util/extract_GO_assignments_from_Trinotate_xls.pl --Trinotate_xls ctena.trinotate.report.xls -G --include_ancestral_terms >GO_annotations.txt

 # Run GO enrichment analysis – create a file containing transcript lengths
$TRINITY_HOME/util/misc/fasta_seq_length.pl Trinity.fasta >Trinity.fasta.seq_lens

 # Run GO enrichment analysis - Use transcript length file to create gene length file:
$TRINITY_HOME/util/misc/TPM_weighted_gene_length.py --gene_trans_map Trinity.fasta.gene_trans_map --trans_length Trinity.fasta.seq_lens --TPM_matrix isoform.TMM.EXPR.matrix >Trinity.gene_lengths.txt

 # Example command to extract TMM-normalized TPM counts of differentially expressed transcripts at a specified p-value cutoff for FDR (-P option; default 0.001) and fold-change (-C option; default 2 = 2^2 = 4-fold) values, and extract depleted and enriched GO terms. Run this command in the differential analysis results folder.

$TRINITY_HOME/Analysis/DifferentialExpression/analyze_diff_expr.pl –matrix ctena.gene.TMM.EXPR.matrix --samples metadata.txt --examine_GO_enrichment --GO_annots ../../ctena_trinity_out/GO_ann
otations.txt --gene_lengths Trinity.gene_lengths.txt

REFERENCES

Abaibou H, Pommier J, Benoit S, Giordano G, Mandrand-Berthelot M (1995). Expression and characterization of the *Escherichia coli fdo* locus and a possible physiological role for aerobic formate dehydrogenase. J Bacteriol 177:7141-7149.

Abanda-Nkpwatt D, Musch M, Tschiersch J, Boettner M, Schwab W (2006). Molecular interaction between *Methylobacterium extorquens* and seedlings: growth promotion, methanol consumption, and localization of the methanol emission site. J Exp Bot 57:4025-4032, doi:10.1093/jxb/erl173.

Ades SE (2004). Control of the alternative sigma factor sigmaE in *Escherichia coli*. Curr Opin Microbiol 7:157-162, doi:10.1016/j.mib.2004.02.010.

Ahmadjian V (1993). The Lichen Symbiosis. New York: John Wiley and Sons.

Alatalo P, Berg Jr. CJ, D'Asaro CN (1984). Reproduction and development in the lucinid clam *Codakia orbicularis* (Linné, 1758). Bulletin of Marine Science 34:424-434.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J Mol Biol 215:403-410, doi:10.1016/S0022-2836(05)80360-2.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol 56:1919-1925.

Anderson LC (2014). Relationships of internal shell features to chemosymbiosis, life position, and geometric constraints within the Lucinidae (bivalvia). In: Experimental Approaches to Understanding Fossil Organisms: Lessons from the Living. Springer, pp 49-72.

Ankrah NY, Luan J, Douglas AE (2017). Cooperative metabolism in a three-partner insect-bacterial symbiosis revealed by metabolic modeling. J Bacteriol, doi:10.1128/JB.00872-16.

Anes J, McCusker MP, Fanning S, Martins M (2015). The ins and outs of RND efflux pumps in *Escherichia coli*. Front Microbiol 6:587, doi:10.3389/fmicb.2015.00587.

Anisimova M, Gascuel O (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 55:539-552, doi:10.1080/10635150600755453.

Antipov D, Korobeynikov A, McLean JS, Pevzner PA (2016). hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32:1009-1015, doi:10.1093/bioinformatics/btv688.

Arp AJ, Childress JJ (1983). Sulfide binding by the blood of the hydrothermal vent tube worm *Riftia pachyptila*. Science 219:295-297, doi:10.1126/science.219.4582.295.

Arp AJ, Childress JJ (1981). Blood function in the hydrothermal vent vestimentiferan tube worm. Science 213:342-344, doi:10.1126/science.213.4505.342.

Azcarate-Peril M, Bruno-Barcena JM, Hassan HM, Klaenhammer TR (2006). Transcriptional and functional analysis of oxalyl-coenzyme A (CoA) decarboxylase and formyl-CoA transferase genes from *Lactobacillus acidophilus*. Appl Environ Microbiol 72:1891-1899, doi:10.1128/AEM.72.3.1891-1899.2006.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA*, et al.* (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75-2164-9-75, doi:10.1186/1471-2164-9-75.

Babraham Bioinformatics (2010). FastQC - a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Babraham Bioinformatics (2007). SeqMonk - A tool to visualise and analyse high throughput mapped sequence data. http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/.

Ball AD, Purdy KJ, Glover EA, Taylor JD (2009). Ctenidial structure and three bacterial symbiont morphotypes in *Anodontia* (*Euanodontia*) ovum (Reeve, 1850) from the Great Barrier Reef, Australia (Bivalvia: Lucinidae). J Molluscan Stud 75:175-185, doi:10.1093/mollus/eyp009.

Barnes PAG (1993). Eco-physiology of the endosymbiont-bearing lucinid bivalve, *Codakia orbiculata*. PhD thesis (University of Plymouth, Plymouth, England), https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.357092

Bartlett MS (1937). Properties of sufficiency and statistical tests. Proc R Soc Lond A 160:268-282, doi:10.1098/rspa.1937.0109.

Beinart RA, Sanders JG, Faure B, Sylva SP, Lee RW, Becker EL, *et al.* (2012). Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. Proc Natl Acad Sci USA 109:E3241-E3250, doi:10.1073/pnas.1202690109.

Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2014). GenBank. Nucleic Acids Res, doi:10.1093/nar/gkt1030.

Bertrand EM, Saito MA, Jeon YJ, Neilan BA (2011). Vitamin B12 biosynthesis gene diversity in the Ross Sea: the identification of a new group of putative polar B12 biosynthesizers. Environ Microbiol 13:1285-1298, doi:10.1111/j.1462-2920.2011.02428.x.

Bianucci E, Furlan A, Castro S (2017). Importance of glutathione in the legume-Rhizobia symbiosis. In: Glutathione in Plant Growth, Development, and Stress Tolerance. Springer International Publishing: Cham, pp 373-396.

Blazejak A, Erseus C, Amann R, Dubilier N (2005). Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru margin. Appl Environ Microbiol 71:1553-1561, doi:10.1128/AEM.71.3.1553-1561.2005.

Bray JR, Curtis JT (1957). An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr 27:326-349, doi:10.2307/1942268.

Borisov VB, Gennis RB, Hemp J, Verkhovsky MI (2011). The cytochrome bd respiratory oxygen reductases. Biochim Biophys Acta 1807:1398-1413, doi:10.1016/j.bbabio.2011.06.016.

Boutet I, Ripp R, Lecompte O, Dossat C, Corre E, Tanguy A, *et al.* (2011). Conjugating effects of symbionts and environmental factors on gene expression in deep-sea hydrothermal vent mussels. BMC Genomics 12:530, doi:10.1186/1471-2164-12-530.

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, *et al.* (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725-731, doi:10.1038/nbt.3893.

Breznak JA, Warnecke F (2008). *Spirochaeta cellobiosiphila* sp. nov., a facultatively anaerobic, marine spirochaete. Int J Syst Evol Microbiol 58:2762-2768, doi:10.1099/ijs.0.2008/001263-0.

Bright M, Bulgheresi S (2010). A complex journey: transmission of microbial symbionts. Nat Rev Microbiol 8:218-230, doi:10.1038/nrmicro2262.

Brissac T, Gros O, Mercot H (2009). Lack of endosymbiont release by two Lucinidae (Bivalvia) of the genus *Codakia*: consequences for symbiotic relationships. FEMS Microbiol Ecol 67:261-267, doi:10.1111/j.1574-6941.2008.00626.x.

Brissac T, Mercot H, Gros O (2011). Lucinidae/sulfur-oxidizing bacteria: ancestral heritage or opportunistic association? Further insights from the Bohol Sea (the Philippines). FEMS Microbiol Ecol 75:63-76, doi:10.1111/j.1574-6941.2010.00989.x.

Brissac T, Higuet D, Gros O, Mercot H (2016). Unexpected structured intraspecific diversity of thioautotrophic bacterial gill endosymbionts within the Lucinidae (Mollusca: Bivalvia). Mar Biol 163:176, doi:10.1007/s00227-016-2949-0.

Brissac T, Rodrigues CF, Gros O, Duperron S (2011). Characterization of bacterial symbioses in *Myrtea* sp. (Bivalvia: Lucinidae) and *Thyasira* sp. (Bivalvia: Thyasiridae) from a cold seep in the Eastern Mediterranean. Mar Ecol 32:198-210, doi:10.1111/j.1439-0485.2010.00413.x.

Brown CT, Olm MR, Thomas BC, Banfield JF (2016). Measurement of bacterial replication rates in microbial communities. Nat Biotechnol 34:1256-1263, doi:10.1038/nbt.3704.

Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T (2011). Bacterial community assembly based on functional genes rather than species. Proc Natl Acad Sci U S A 108:14288-14293, doi:10.1073/pnas.1101591108.

Campbell BJ, Cary SC (2001). Characterization of a novel spirochete associated with the hydrothermal vent polychaete annelid, *Alvinella pompejana*. Appl Environ Microbiol 67:110-117, doi:10.1128/AEM.67.1.110-117.2001.

Carlstrom CI, Loutey DE, Wang O, Engelbrektson A, Clark I, Lucas LN*, et al.* (2015). Phenotypic and genotypic description of *Sedimenticola selenatireducens* strain CUZ, a marine (per)chlorate-respiring gammaproteobacterium, and its close relative the chlorate-respiring *Sedimenticola* strain NSS. Appl Environ Microbiol 81:2717-2726, doi:10.1128/AEM.03606-14.

Carney SL, Flores JF, Orobona KM, Butterfield DA, Fisher CR, Schaeffer SW (2007). Environmental differences in hemoglobin gene expression in the hydrothermal vent tubeworm, *Ridgeia piscesae*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 146:326-337, doi:10.1016/j.cbpb.2006.11.002.

Caro A, Gros O, Got P, De Wit R, Troussellier M (2007). Characterization of the population of the sulfur-oxidizing symbiont of *Codakia orbicularis* (Bivalvia, Lucinidae) by single-cell analyses. Appl Environ Microbiol 73:2101-9, doi:10.1128/AEM.01683-06.

Caro A, Got P, Bouvy M, Troussellier M, Gros O (2009). Effects of long-term starvation on a host bivalve (*Codakia orbicularis*, Lucinidae) and its symbiont population. Appl Environ Microbiol 75:3304-13, doi:10.1128/AEM.02659-08.

Cary SC, Vetter RD, Felbeck H (1989). Habitat characterization and nutritional strategies of the endosymbiont-bearing bivalve *Lucinoma aequizonata*. Mar Ecol Prog Ser 55:31-45.

Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, Postle K, *et al.* (2007). Colicin biology. Microbiol Mol Biol Rev 71:158-229, doi:10.1128/MMBR.00036-06.

Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB (1981). Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: Possible chemoautotrophic symbionts. Science 213:340-342, doi:10.1126/science.213.4505.340.

Cavanaugh CM (1983). Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats. Nature 302:58-61, doi:10.1038/302058a0.

Cavanaugh CM, McKiness ZP, Newton ILG, Stewart FJ (2006). Marine chemosynthetic symbioses. In: Prokaryotes. Springer: New York, pp 475-507.

Chen NH, Djoko KY, Veyrier FJ, McEwan AG (2016). Formaldehyde stress responses in bacterial pathogens. Front Microbiol 7:257, doi:10.3389/fmicb.2016.00257.

Chipman D, Barak Z, Schloss JV (1998). Biosynthesis of 2-aceto-2-hydroxy acids: acetolactate synthases and acetohydroxyacid synthases. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1385:401-419, doi:10.1016/S0167-4838(98)00083-1.

Choi EJ, Kwon HC, Sohn YC, Yang HO (2010). *Kistimonas asteriae* gen. nov., sp. nov., a gammaproteobacterium isolated from *Asterias amurensis*. Int J Syst Evol Microbiol 60:938-943, doi:10.1099/ijs.0.014282-0.

Christo SW, Ivachuk CS, Ferreira-Junior AL, Absher TM (2016). Reproductive periods of *Lucina pectinata* (Bivalve; Lucinidae) in the Paranagua Estuarine Complex, Parana - Brazil. Braz J Biol 76:300-306, doi:10.1590/1519-6984.12514.

Cooper G (2000). The mechanism of vesicular transport. In: The Cell: A Molecular Approach. Sinauer Associates: Sunderland (MA).

Cort JR, Selan U, Schulte A, Grimm F, Kennedy MA, Dahl C (2008). *Allochromatium vinosum* DsrC: solution-state NMR structure, redox properties, and interaction with DsrEFH, a protein essential for purple sulfur bacterial sulfur oxidation. J Mol Biol 382:692-707, doi:10.1016/j.jmb.2008.07.022.

Croes LM, Meijer WG, Dijkhuizen L (1991). Regulation of methanol oxidation and carbon dioxide fixation in *Xanthobacter* strain 25a grown in continuous culture. Arch Microbiol 155:159-163, doi:10.1007/BF00248611.

Crump BC, Wojahn JM, Tomas F, Mueller RS (2018). Metatranscriptomics and amplicon sequencing reveal mutualisms in seagrass microbiomes. Front Microbiol 9:388, doi:10.3389/fmicb.2018.00388.

Dando PR, Southward AJ, Southward EC, Terwilliger NB, Terwilliger RC (1985). Sulphur-oxidizing bacteria and haemoglobin in gills of the bivalve mollusc *Myrtea spinifera*. Mar Ecol Prog Ser 23:85-98.

Dando PR, Southward AJ, Southward EC (1986). Chemoautotrophic symbionts in the gills of the bivalve mollusc *Lucinoma borealis* and the sediment chemistry of its habitat. Proc R Soc Lond [Biol] 227:227-247.

Dando PR, Ridgway SA, Spiro B (1994). Sulphide 'mining' by lucinid bivalve molluscs: demonstrated by stable sulphur isotope measurements and experimental models. Mar Eco Prog Ser 107:169-175.

Davidson SK, Koropatnick TA, Kossmehl R, Sycuro L, McFall-Ngai MJ (2004). NO means 'yes' in the squid-vibrio symbiosis: nitric oxide (NO) during the initial stages of a beneficial association. Cell Microbiol 6:1139-1151, doi:10.1111/j.1462-5822.2004.00429.x.

de Almeida A, Nikel PI, Giordano AM, Pettinari MJ (2007). Effects of granule-associated protein PhaP on glycerol-dependent growth and polymer production in poly(3-hydroxybutyrate)-producing *Escherichia coli*. Appl Environ Microbiol 73:7912-7916, doi:10.1128/AEM.01900-07.

Decker C, Olu K, Arnaud-Haond S, Duperron S (2013). Physical proximity may promote lateral acquisition of bacterial symbionts in vesicomyid clams. PLoS One 8:e64830, doi:10.1371/journal.pone.0064830.

Delepelaire P (2004). Type I secretion in gram-negative bacteria. Biochim Biophys Acta 1694:149-161, doi:10.1016/j.bbamcr.2004.05.001.

DeLong EF, Wickham GS, Pace NR (1989). Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. Science 243:1360-1363, doi:10.1126/science.2466341.

Dibrova DV, Cherepanov DA, Galperin MY, Skulachev VP, Mulkidjanian AY (2013). Evolution of cytochrome bc complexes: from membrane-anchored dehydrogenases of ancient bacteria to triggers of apoptosis in vertebrates. Biochim Biophys Acta 1827:1407-1427, doi:10.1016/j.bbabio.2013.07.006.

Diepold A, Armitage JP (2015). Type III secretion systems: the bacterial flagellum and the injectisome. Philos Trans R Soc Lond B Biol Sci 370:20150020, doi:10.1098/rstb.2015.0020.

Dincturk HB, Demir V, Aykanat T (2011). Bd oxidase homologue of photosynthetic purple sulfur bacterium *Allochromatium vinosum* is co-transcribed with a nitrogen fixation related gene. Antonie Van Leeuwenhoek 99:211-220, doi:10.1007/s10482-010-9478-5.

Ding JY, Shiu JH, Chen WM, Chiang YR, Tang SL (2016). Genomic insight into the host-endosymbiont relationship of *Endozoicomonas montiporae* CL-33(T) with its coral host. Front Microbiol 7:251, doi:10.3389/fmicb.2016.00251.

Dirks RM, Pierce NA (2004). Triggered amplification by hybridization chain reaction. Proc Natl Acad Sci USA 101:15275-15278, doi:10.1073/pnas.0407024101.

Distel DL, Felbeck H (1987). Endosymbiosis in the lucinid clams *Lucinoma aequizonata*, *Lucinoma annulata* and *Lucina floridana*: a re-examination of the functional morphology of the gills as bacteria-bearing organs. Mar Biol 96:79-86, doi:10.1007/BF00394840.

Distel DL, Lee HK, Cavanaugh CM (1995). Intracellular coexistence of methano- and thioautotrophic bacteria in a hydrothermal vent mussel. Proc Natl Acad Sci USA 92:9598-9602.

Distel DL, Altamia MA, Lin Z, Shipway JR, Han A, Forteza I*, et al.* (2017). Discovery of chemoautotrophic symbiosis in the giant shipworm *Kuphus polythalamia* (Bivalvia: Teredinidae) extends wooden-steps theory. Proc Natl Acad Sci U S A 114:E3652-E3658, doi:10.1073/pnas.1620470114.

Dmytrenko O, Russell SL, Loo WT, Fontanez KM, Liao L, Roeselers G*, et al.* (2014). The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. BMC Genomics 15:924, doi:10.1186/1471-2164-15-924.

Dolan SK, Wijaya A, Geddis SM, Spring DR, Silva-Rocha R, Welch M (2018). Loving the poison: the methylcitrate cycle and bacterial pathogenesis. Microbiology 164:251-259, doi:10.1099/mic.0.000604.

Dong X, Greening C, Bruls T, Conrad R, Guo K, Blaskowski S, *et al.* (2018). Fermentative Spirochaetes mediate necromass recycling in anoxic hydrocarbon-contaminated habitats. ISME J 12:2039-2050, doi:10.1038/s41396-018-0148-3.

Doty TW (2015). Environmental controls on the diversity and distribution of endosymbionts associated with *Phacoides pectinatus* (Bivalvia: Lucinidae) from shallow mangrove and seagrass sediments, St. Lucie County, Florida. MSc thesis (University of Tennessee, Knoxville, Tennessee, USA), http://trace.tennessee.edu/utk_gradthes/3548/.

Dubilier N, Bergin C, Lott C (2008). Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Microbiol 6:725-740, doi:10.1038/nrmicro1992.

Dubilier N, Amann R, Erseus C, Muyzer G, Park S, Giere O*, et al.* (1999). Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete *Olavius loisae* (Annelida). Mar Ecol Prog Ser 178:271-280.

Duperron S, Nadalig T, Caprais JC, Sibuet M, Fiala-Medioni A, Amann R*, et al.* (2005). Dual symbiosis in a *Bathymodiolus* sp. mussel from a methane seep on the Gabon continental margin (Southeast Atlantic): 16S rRNA phylogeny and distribution of the symbionts in gills. Appl Environ Microbiol 71:1694-1700, doi:10.1128/AEM.71.4.1694-1700.2005.

Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP*, et al.* (2006). A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. Environ Microbiol 8:1441-1447, doi:10.1111/j.1462-2920.2006.01038.x.

Duperron S, Fiala-Medioni A, Caprais J, Olu K, Sibuet M (2007). Evidence for chemoautotrophic symbiosis in a Mediterranean cold seep clam (Bivalvia: Lucinidae): comparative sequence analysis of bacterial 16S rRNA, APS reductase and RuBisCO genes. FEMS Microbiol Ecol 59:64-70, doi:10.1111/j.1574-6941.2006.00194.x.

Duperron S, Rodrigues CF, Leger N, Szafranski K, Decker C, Olu K*, et al.* (2012). Diversity of symbioses between chemosynthetic bacteria and metazoans at the Guiness cold seep site (Gulf of Guinea, West Africa). MicrobiologyOpen 1:467-480.

Duplessis MR, Dufour SC, Blankenship LE, Felbeck H, Yayanos AA (2004). Anatomical and experimental evidence for particulate feeding in *Lucinoma aequizonata* and *Parvilucina tenuisculpta* (Bivalvia : Lucinidae) from the Santa Barbara Basin. Mar Biol 145:551-561, doi:10.1007/s00227-004-1350-6.

Duplessis MR, Ziebis W, Gros O, Caro A, Robidart J, Felbeck H (2004). Respiration strategies utilized by the gill endosymbiont from the host lucinid *Codakia orbicularis* (Bivalvia: Lucinidae). Appl Environ Microbiol 70:4144-4150, doi:10.1128/AEM.70.7.4144-4150.2004.

Durand P, Gros O (1996). Bacterial host specificity of Lucinacea endosymbionts: interspecific variation in 16S rRNA sequences. FEMS Microbiol Lett 140:193-8, doi:10.1016/0378-1097(96)00178-4.

Durand P, Gros O, Frenkiel L, Prieur D (1996). Phylogenetic characterization of sulfur-oxidizing bacterial endosymbionts in three tropical Lucinidae by 16S rDNA sequence analysis. Mol Mar Biol Biotech 5:37-42.

Eddie BJ, Hanson TE (2013). *Chlorobaculum tepidum* TLS displays a complex transcriptional response to sulfide addition. J Bacteriol 195:399-408, doi:10.1128/JB.01342-12.

Edgar RC (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113, doi:10.1186/1471-2105-5-113.

Edgar RC (2018). Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics , doi:10.1093/bioinformatics/bty113.

Ehara T, Kitajima S, Kanzawa N, Tamiya T, Tsuchiya T (2002). Antimicrobial action of achacin is mediated by L-amino acid oxidase activity. FEBS Lett 531:509-512, doi:10.1016/S0014-5793(02)03608-6.

Felbeck H (1981). Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). Science 213:336-338, doi:10.1126/science.213.4505.336.

Felbeck H, Childress JJ, Somero GN (1981). Calvin-Benson cycle and sulphide oxidation enzymes in animals from sulphide-rich habitats. Nature 293:291-293, doi:10.1038/293291a0.

Ferdy J, Godelle B (2005). Diversification of transmission modes and the evolution of mutualism. Am Nat 166:613-627, doi:10.1086/491799.

Fisher MR, Hand SC (1984). Chemoautotrophic symbionts in the bivalve *Lucina floridana* from seagrass beds. Biol Bull 167:445-459, doi:10.2307/1541289.

Fligner MA, Killeen TJ (1976). Distribution-free two-sample tests for scale. Journal of the American Statistical Association 71:210-213, doi:10.2307/2285771.

Flood BE, Jones DS, Bailey JV (2015). *Sedimenticola thiotaurini* sp. nov., a sulfur-oxidizing bacterium isolated from salt marsh sediments, and emended descriptions of the genus *Sedimenticola* and *Sedimenticola selenatireducens*. Int J Syst Evol Microbiol 65:2522-2530, doi:10.1099/ijs.0.000295.

Frenkiel L, Gros O, Mouëza M (1997). Storage tissue and reproductive strategy in *Lucina pectinata* (Gmelin), a tropical lucinid bivalve adapted to a reducing sulfur-rich, mangrove environment. Invertebr Reprod Dev 31:199-210, doi:10.1080/07924259.1997.9672577.

Frenkiel L, Gros O, Mouëza M (1996). Gill structure in *Lucina pectinata* (Bivalvia: Lucinidae) with reference to hemoglobin in bivalves with symbiotic sulphur-oxidizing bacteria. Mar Biol 125:511-524, doi:10.1007/BF00353264.

Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J (2001). Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? Appl Environ Microbiol 67:2873-2882, doi:10.1128/AEM.67.7.2873-2882.2001.

Fu L, Niu B, Zhu Z, Wu S, Li W (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150-3152, doi:10.1093/bioinformatics/bts565.

Garcia-Horsman JA, Barquera B, Rumbley J, Ma J, Gennis RB (1994). The superfamily of heme-copper respiratory oxidases. J Bacteriol 176:5587-5600.

Garcia JR, Gerardo NM (2014). The symbiont side of symbiosis: do microbes really benefit? Front Microbiol 5:510, doi:10.3389/fmicb.2014.00510.

Garcias-Bonet N, Duarte CM (2017). Methane production by seagrass ecosystems in the Red Sea. Front Mar Sci 4:340, doi:10.3389/fmars.2017.00340.

Gardebrecht A, Markert S, Sievert SM, Felbeck H, Thurmer A, Albrecht D*, et al.* (2012). Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. ISME J 6:766-776, doi:10.1038/ismej.2011.137.

Gene Ontology Consortium (2015). Gene ontology consortium: Going forward. Nucleic Acids Res 43:D1049-D1056, doi:10.1093/nar/gku1179.

Genkai-Kato M, Yamamura N (1999). Evolution of mutualistic symbiosis without vertical transmission. Theor Popul Biol 55:309-323, doi:10.1006/tpbi.1998.1407.

Ghosh W, Dam B (2009). Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. FEMS Microbiol Rev 33:999-1043, doi:10.1111/j.1574-6976.2009.00187.x.

Girguis PR, Childress JJ (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, doi:10.1242/jeb.02404.

Glover EA, Taylor JD, Rowden AA (2004). *Bathyaustriella thionipta*, a new lucinid bivalve from a hydrothermal vent on the Kermadec Ridge, New Zealand and its relationship to shallow-water taxa (Bivalvia: Lucinidae). J Molluscan Stud 70:283-295, doi:10.1093/mollus/70.3.283.

Glover EA, Taylor JD, Williams ST (2008). Mangrove associated lucinid bivalves of the central Indo-West Pacific: review of the "*Austriella*" group with a new genus and species (Mollusca: Bivalvia: Lucinidae). The Raffles Bulletin of Zoology S18:25-40.

Goemann AM (2015). Rare occurrences of free-living bacteria belonging to *Sedimenticola* from subtidal seagrass beds associated with the lucinid clam, *Stewartia floridana*. MSc thesis (University of Tennessee, Knoxville, Tennessee, USA), http://trace.tennessee.edu/utk_gradthes/3549/.

Goffredi SK, Yi H, Zhang Q, Klann JE, Struve IA, Vrijenhoek RC*, et al.* (2014). Genomic versatility and functional variation between two dominant heterotrophic symbionts of deep-sea *Osedax* worms. ISME J 8:908-924, doi:10.1038/ismej.2013.201.

Good IJ (1953). The population frequencies of species and the estimation of population parameters. Biometrika 40:237-264.

Goto S, Kato S, Kimura T, Muto A, Himeno H (2011). RsgA releases RbfA from 30S ribosome during a late stage of ribosome biosynthesis. EMBO J 30:104-114, doi:10.1038/emboj.2010.291.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I*, et al.* (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol 29:644-652, doi:10.1038/nbt.1883.

Green ER, Mecsas J (2016). Bacterial secretion systems: an overview. In: Virulence Mechanisms of Bacterial Pathogens. ASM Press: Washington, DC, pp 215-239.

Green-García AM, Engel AS (2012). Bacterial diversity of siliciclastic sediments in a *Thalassia testudinum* meadow and the implications for *Lucinisca nassula* chemosymbiosis. Estuar Coast Shelf Sci 112:153-161, doi:10.1016/j.ecss.2012.07.010.

Green-García AM (2008). Characterization of the lucinid bivalve-bacteria symbiotic system: The significance of the geochemical habitat on bacterial symbiont diversity and phylogeny. MSc thesis, (Louisiana State University, Baton Rouge, Louisiana, USA), https://digitalcommons.lsu.edu/gradschool_theses/1970/.

Gros O, Frenkiel L, Mouëza M (1996a). Gill ultrastructure and symbiotic bacteria in the tropical lucinid, *Linga pensylvanica* (Linne). Symbiosis 20:259-280.

Gros O, Darrasse A, Durand P, Frenkiel L, Mouëza M (1996b). Environmental transmission of a sulfur-oxidizing bacterial gill endosymbiont in the tropical lucinid bivalve *Codakia orbicularis*. Appl Environ Microbiol 62:2324-30.

Gros O, Frenkiel L, Mouëza M (1997). Embryonic, larval, and post-larval development in the symbiotic clam *Codakia orbicularis* (Bivalvia: Lucinidae). Invertebr Biol 116:86-101, doi:10.2307/3226973.

Gros O, De Wulf-Durand P, Frenkiel L, Mouëza M (1998). Putative environmental transmission of sulfur-oxidizing bacterial symbionts in tropical lucinid bivalves inhabiting various environments. FEMS Microbiol Lett 160:257-262, doi:10.1016/S0378-1097(98)00041-X.

Gros O, Duplessis MR, Felbeck H (1999). Embryonic development and endosymbiont transmission mode in the symbiotic clam *Lucinoma aequizonata* (Bivalvia: Lucinidae). Invertebr Reprod Dev 36:93-103.

Gros O, Liberge M, Heddi A, Khatchadourian C, Felbeck H (2003a). Detection of the free-living forms of sulfide-oxidizing gill endosymbionts in the lucinid habitat (*Thalassia testudinum* environment). Appl Environ Microbiol 69:6264-7, doi:10.1128/AEM.69.10.6264-6267.2003.

Gros O, Liberge M, Felbeck H (2003b). Interspecific infection of aposymbiotic juveniles of Codakia orbicularis by various tropical lucinid gill-endosymbionts. Mar Biol 142:57-66, doi:10.1007/s00227-002-0921-7.

Gros O, Elisabeth NH, Gustave SD, Caro A, Dubilier N (2012). Plasticity of symbiont acquisition throughout the life cycle of the shallow-water tropical lucinid *Codakia orbiculata* (Mollusca: Bivalvia). Environ Microbiol 14:1584-1595, doi:10.1111/j.1462-2920.2012.02748.x.

Gunderson FF, Cianciotto NP (2013). The CRISPR-associated gene *cas2* of *Legionella pneumophila* is required for intracellular infection of amoebae. MBio 4:e00074-13, doi:10.1128/mBio.00074-13.

Guo C, Liu S, Yao Y, Zhang Q, Sun MZ (2012). Past decade study of snake venom L-amino acid oxidase. Toxicon 60:302-311, doi:10.1016/j.toxicon.2012.05.001.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075, doi:10.1093/bioinformatics/btt086.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J*, et al.* (2013). *De novo* transcript sequence reconstruction from RNA-Seq: Reference generation and analysis with Trinity. Nature protocols 8:1494-1512, doi:10.1038/nprot.2013.084.

Hahn MW, Stadler P, Wu QL, Pöckl M (2004). The filtration-acclimatization method for isolation of an important fraction of the not readily cultivable bacteria. J Microbiol Methods 57:379-390, doi:10.1016/j.mimet.2004.02.004.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95-98.

Hansen AK, Moran NA (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc Natl Acad Sci USA 108:2849-2854, doi:10.1073/pnas.1013465108.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R*, et al.* (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258-61, doi:10.1093/nar/gkh036.

Harwood CS, Canale-Parola E (1984). Ecology of spirochetes. Annu Rev Microbiol 38:161-192, doi:10.1146/annurev.mi.38.100184.001113.

Hasegawa M, Kishino H, Yano T (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160-174.

Hemrajani C, Berger CN, Robinson KS, Marches O, Mousnier A, Frankel G (2010). NleH effectors interact with Bax inhibitor-1 to block apoptosis during enteropathogenic *Escherichia coli i*nfection. Proc Natl Acad Sci U S A 107:3129-3134, doi:10.1073/pnas.0911609106.

Hentschel U, Cary SC, Felbeck H (1993). Nitrate respiration in chemoautotrophic symbionts of the bivalve *Lucinoma aequizonata*. Mar Ecol Prog Ser 94:35-41, doi:10.3354/meps094035.

Hentschel U, Felbeck H (1995). Nitrate respiration in chemoautotrophic symbionts of the bivalve *Lucinoma aequizonata* is not regulated by oxygen. Appl Environ Microbiol 61:1630-1633.

Hentschel U, Hand S, Felbeck H (1996). The contribution of nitrate respiration to the energy budget of the symbiont-containing clam *Lucinoma aequizonata*: a calorimetric study. J Exp Biol 199:427-433.

Herry A, Diouris M, 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, doi:10.1007/BF00428126.

Higgs ND, Newton J, Attrill MJ (2016). Caribbean spiny lobster fishery is underpinned by trophic subsidies from chemosynthetic primary production. Curr Biol 26:3393-3398, doi:10.1016/j.cub.2016.10.034.

Husmann G, Gerdts G, Wichels A (2010). Spirochetes in crystalline styles of marine bivalves: group-specific PCR detection and 16S rRNA sequence analysis. Journal of Shellfish Research 29:1069-1075.

Huson DH, Auch AF, Qi J, Schuster SC (2007). MEGAN analysis of metagenomic data. Genome Res 17:377-386, doi:10.1101/gr.5969107.

Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S*, et al.* (2016). Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. ISME J 10:990-1001, doi:10.1038/ismej.2015.176.

Jensen SI, Kuhl M, Glud RN, Jorgensen LB, Prieme A (2005). Oxic microzones and radial oxygen loss from roots of *Zostera marina*. Mar Ecol Prog Ser 293:49-58, doi:10.3354/meps293049.

Jensen S, Duperron S, Birkeland NK, Hovland M (2010). Intracellular Oceanospirillales bacteria inhabit gills of *Acesta* bivalves. FEMS Microbiol Ecol 74:523-533, doi:10.1111/j.1574-6941.2010.00981.x.

Johnson MA, Fernandez C (2001). Bacterial symbiosis in *Loripes lucinalis* (Mollusca : Bivalvia) with comments on reproductive strategy. J Mar Biol Ass U K 81:251-257.

Johnson MA, Fernandez C, Pergent G (2002). The ecological importance of an invertebrate chemoautotrophic symbiosis to phanerogam seagrass beds. Bulletin of Marine Science 71:1343-1351.

Jones DS, Quitmyer IR (1996). Marking time with bivalve shells: oxygen isotopes and season of annual increment formation. Palaios 11:340-346, doi:10.2307/3515244.

Kalmar Band Greensmith L (2009). Induction of heat shock proteins for protection against oxidative stress. Advanced Drug Delivery Reviews 61:310-318, doi:10.1016/j.addr.2009.02.003.

Kaminski PA, Kitts CL, Zimmerman Z, Ludwig RA (1996). *Azorhizobium caulinodans* uses both cytochrome bd (quinol) and cytochrome cbb3 (cytochrome c) terminal oxidases for symbiotic N2 fixation. J Bacteriol 178:5989-5994.

Kang DD, Froula J, Egan R, Wang Z (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165, doi:10.7717/peerj.1165.

Katharios P, Seth-Smith HM, Fehr A, Mateos JM, Qi W, Richter D*, et al.* (2015). Environmental marine pathogen isolation using mesocosm culture of sharpsnout seabream: striking genomic and morphological features of novel *Endozoicomonas* sp. Sci Rep 5:17609, doi:10.1038/srep17609.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, *et al.* (2012). Geneious basic: an integrated and extendable desktop software platform for the

organization and analysis of sequence data. Bioinformatics 28:1647-1649, doi:10.1093/bioinformatics/bts199.

Keppler F, Hamilton JTG, Brab M, Rockmann T (2006). Methane emissions from terrestrial plants under aerobic conditions. Nature 439:187-191, doi:10.1038/nature04420.

Kimura M (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111-120.

Kleiner M, Wentrup C, Holler T, Lavik G, Harder J, Lott C, *et al.* (2015). Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments. Environ Microbiol 17:5023-5035, doi:10.1111/1462-2920.12912.

Kleiner M, Petersen JM., Dubilier N (2012). Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. Curr Opin Microbiol 15:621-31, doi:10.1016/j.mib.2012.09.003.

Knight JM, Griffin L, Dale PER, Sheaves M (2013). Short-term dissolved oxygen patterns in sub-tropical mangroves. Estuar Coast Shelf Sci 131:290-296, doi:10.1016/j.ecss.2013.06.024.

Kobayashi K, Ogata H, Morikawa M, Iijima S, Harada N, Yoshida T*, et al.* (2002). Distribution and partial characterisation of IgG Fc binding protein in various mucin producing cells and body fluids. Gut 51:169-176.

König S, Gros O, Heiden SE, Hinzke T, Thurmer A, Poehlein A*, et al.* (2016). Nitrogen fixation in a chemoautotrophic lucinid symbiosis. Nat Microbiol 2:16193, doi:10.1038/nmicrobiol.2016.193.

Konstantinidis KT, Tiedje JM (2005). Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258-6264, doi:10.1128/JB.187.18.6258-6264.2005.

Kopylova E, Noe L, Touzet H (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211-3217, doi:10.1093/bioinformatics/bts611.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112-5120, doi:10.1128/AEM.01043-13.

Kraus DW, Wittenberg JB (1990). Hemoglobins of the *Lucina pectinata*/bacteria symbiosis. I. Molecular properties, kinetics and equilibria of reactions with ligands. J Biol Chem 27:16043-16053.

Kruskal WH, Wallis WA (1952). Use of ranks in one-criterion variance analysis. JASA 47:583-621, doi:10.2307/2280779.

Kumar S, Stecher G, Tamura K (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870-1874, doi:10.1093/molbev/msw054.

Kurahashi M, Yokota A (2007). *Endozoicomonas elysicola* gen. nov., sp. nov., a gamma-proteobacterium isolated from the sea slug *Elysia ornata*. Syst Appl Microbiol 30:202-206, doi:10.1016/j.syapm.2006.07.003.

Lane DJ (1991). 16S/23S rRNA sequencing. In: Nucleic Acid Techniques in Bacterial Systematics. John Wiley and Sons: New York, pp 115-175.

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nature methods 9:357-359, doi:10.1038/nmeth.1923.

Law CW, Chen Y, Shi W, Smyth GK (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 15:R29, doi:10.1186/gb-2014-15-2-r29.

Le Pennec M, Beninger PG, Herry A (1995). Feeding and digestive adaptations of bivalve molluscs to sulphide-rich habitats. Comp Biochem Physiol 111:183-189, doi:10.1016/0300-9629(94)00211-B.

Le Pennec M, Beninger PG (2000). Reproductive characteristics and strategies of reducing-system bivalves. Comparative Biochemistry and Physiology 126:1-16, doi:10.1016/S0742-8413(00)00100-6.

Le SQ, Gascuel O (2008). An improved general amino acid replacement matrix. Mol Biol Evol 25:1307-1320, doi:10.1093/molbev/msn067.

Lee J, Shin NR, Lee HW, Roh SW, Kim MS, Kim YO*, et al.* (2012). *Kistimonas scapharcae* sp. nov., isolated from a dead ark clam (*Scapharca broughtonii*), and emended description of the genus *Kistimonas*. Int J Syst Evol Microbiol 62:2865-2869, doi:10.1099/ijs.0.038422-0.

Lemke JJ, Sanchez-Vazquez P, Burgos HL, Hedberg G, Ross W, Gourse RL (2011). Direct regulation of *Escherichia coli* ribosomal protein promoters by the transcription factors ppGpp and DksA. Proc Natl Acad Sci U S A 108:5712-5717, doi:10.1073/pnas.1019383108.

Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN (2010). RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26:493-500, doi:10.1093/bioinformatics/btp692.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N*, et al.* (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079, doi:10.1093/bioinformatics/btp352.

Liao Y, Smyth GK, Shi W (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923-930, doi:10.1093/bioinformatics/btt656.

Liberge M, Gros O, Frenkiel L (2001). Lysosomes and sulfide-oxidizing bodies in the bacteriocytes of *Lucina pectinata*, a cytochemical and microanalysis approach. Mar Biol 139:401-409, doi:10.1007/s002270000526.

Liljedahl L (1992). The Silurian *Ilionia prisca*, oldest known deep-burrowing suspension feeding bivalve. J Paleontol 66:206-210.

Login FH, Balmand S, Vallier A, Vincent-Monegat C, Vigneron A, Weiss-Gayet M., *et al.* (2011). Antimicrobial peptides keep insect endosymbionts under control. Science 334:362–365, doi: 10.1126/science.1209728.

Long, BL (2016). Geometric morphometric analyses of environment related shell variation in *Stewartia floridana* (Bivalva: Lucinidae). MSc thesis (South Dakota School of Mines and Technology, Rapid City, South Dakota, USA).

Louie TS, Giovannelli D, Yee N, Narasingarao P, Starovoytov V, Goker M*, et al.* (2016). High-quality draft genome sequence of *Sedimenticola selenatireducens* strain AK4OH1(T), a gammaproteobacterium isolated from estuarine sediment. Stand Genomic Sci 11:66, doi:10.1186/s40793-016-0191-5.

Louwen R, Staals RH, Endtz HP, van Baarlen P, van der Oost J (2014). The role of CRISPR-Cas systems in virulence of pathogenic bacteria. Microbiol Mol Biol Rev 78:74-88, doi:10.1128/MMBR.00039-13.

Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550, doi:10.1186/s13059-014-0550-8.

Maezawa K, Shigenobu S, Taniguchi H, Kubo T, Aizawa S, Morioka M (2006). Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium *Buchnera aphidicola* sp. strain APS. J Bacteriol 188:6539-6543, doi:10.1128/JB.00561-06.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P*, et al.* (2011). Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9:467-477, doi:10.1038/nrmicro2577.

Maquat LE, Tarn, W, Isken, O (2010). The pioneer round of translation: features and functions. Cell 142:368-374, doi: 10.1016/j.cell.2010.07.022.

Marcia M, Ermler U, Peng G, Michel H (2010). A new structure-based classification of sulfide:quinone oxidoreductases. Proteins 78:1073-1083, doi:10.1002/prot.22665.

Marie B, Le Roy N, Zanella-Cleon I, Becchi M, Marin F (2011). Molecular evolution of mollusc shell proteins: insights from proteomic analysis of the edible mussel *Mytilus*. J Mol Evol 72:531-546, doi:10.1007/s00239-011-9451-6.

Markert S, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D*, et al.* (2011). Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. Proteomics 11:3106-3117, doi:10.1002/pmic.201100059.

Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hugler M*, et al.* (2007). Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. Science 315:247-250, doi:10.1126/science.1132913.

Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal 17:10-12, doi:10.14806/ej.17.1.200.

Mausz M, Schmitz-Esser S, Steiner G (2010). Identification and comparative analysis of the endosymbionts of *Loripes lacteus* and *Anodontia fragilis* (Bivalvia: Lucinidae). Unpublished, NCBI accession numbers GQ853555- GQ853556.

McMurdie PJ, Holmes S (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217, doi:10.1371/journal.pone.0061217.

Meier C, Carter LG, Winter G, Owens RJ, Stuart DI, Esnouf RM (2007). Structure of 5-formyltetrahydrofolate cyclo-ligase from *Bacillus anthracis* (BA4489). Acta Crystallogr Sect F Struct Biol Cryst Commun 63:168-172, doi:10.1107/S1744309107007221.

Mendoza M, Guiza L, Martinez X, Caraballo X, Rojas J, Aranguren LF*, et al.* (2013). A novel agent (*Endozoicomonas elysicola*) responsible for epitheliocystis in cobia *Rachycentrum canadum* larvae. Dis Aquat Organ 106:31-37, doi:10.3354/dao02636.

Meyer EL, Nilkerd B, Glover EA, Taylor JD (2008). Ecological importance of chemoautotrophic lucinid bivalves in a peri–mangrove community in Eastern Thailand. RBZ 18:41-55.

Millikan DS, Ruby EG (2004). *Vibrio fischeri* flagellin A is essential for normal motility and for symbiotic competence during initial squid light organ colonization. J Bacteriol 186:4315-4325, doi:10.1128/JB.186.13.4315-4325.2004.

Moore S, Warren M (2012). The anaerobic biosynthesis of vitamin B_{12}. Biochem Soc Trans 40:581-586, doi:10.1042/BST20120066.

Munoz-Elias EJ, Upton AM, Cherian J, McKinney JD (2006). Role of the methylcitrate cycle in *Mycobacterium tuberculosis* metabolism, intracellular growth, and virulence. Mol Microbiol 60:1109-1122, doi:10.1111/j.1365-2958.2006.05155.x.

Munoz-Elias EJ, McKinney JD (2005). *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. Nat Med 11:638-644, doi:10.1038/nm1252.

Mushegian AA and Ebert D (2015). Rethinking "mutualism" in diverse host-symbiont communities. BioEssays 38:100-108, doi:10.1002/bies.201500074.

Nakagawa S, Shimamura S, Takaki Y, Suzuki Y, Murakami S, Watanabe T*, et al.* (2014). Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. ISME J 8:40-51, doi:10.1038/ismej.2013.131.

Narasingarao P, Haggblom MM (2006). *Sedimenticola selenatireducens*, gen. nov., sp. nov., an anaerobic selenate-respiring bacterium isolated from estuarine sediment. Syst Appl Microbiol 29:382-388, doi:10.1016/j.syapm.2005.12.011.

NCBI Resource Coordinators (2016). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44:D7-19, doi:10.1093/nar/gkv1290.

Neave MJ, Michell CT, Apprill A, Voolstra CR (2017). *Endozoicomonas* genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. Sci Rep 7:40579, doi:10.1038/srep40579.

Neave MJ, Apprill A, Ferrier-Pages C, Voolstra CR (2016a). Diversity and function of prevalent symbiotic marine bacteria in the genus *Endozoicomonas*. Appl Microbiol Biotechnol 100:8315-8324, doi:10.1007/s00253-016-7777-0.

Neave MJ, Rachmawati R, Xun L, Michell CT, Bourne DG, Apprill A*, et al.* (2016b). Differential specificity between closely related corals and abundant *Endozoicomonas* endosymbionts across global scales. ISME J, doi:10.1038/ismej.2016.95.

Nemecek-Marshall M, MacDonald RC, Franzen JJ, Wojciechowski CL, Fall R (1995). Methanol emission from leaves (enzymatic detection of gas-phase methanol and relation

of methanol fluxes to stomatal conductance and leaf development). Plant Physiol 108:1359-1368.

Nguyen N, Warnow T, Pop M, White B (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. Npj Biofilms and Microbiomes 2:16004.

Nishiguchi MK (2000). Temperature affects species distribution in symbiotic populations of *Vibrio* spp. Applied and Environmental Microbiology 66:3550-3555, doi:10.1128/AEM.66.8.3550-3555.2000.

Nivaskumar M, Francetic O (2014). Type II secretion system: a magic beanstalk or a protein escalator. Biochim Biophys Acta 1843:1568-1577, doi:10.1016/j.bbamcr.2013.12.020.

Nonaka G, Blankschien M, Herman C, Gross CA, Rhodius VA (2006). Regulon and promoter analysis of the *E. coli* heat-shock factor, sigma32, reveals a multifaceted cellular response to heat stress. Genes Dev 20:1776-1789, doi:0.1101/gad.1428206.

Nyholm SV, McFall-Ngai M (2003). Dominance of *Vibrio fischeri* in secreted mucus outside the light organ of *Euprymna scolopes*: the first site of symbiont specificity. Appl Environ Microbiol 69:3932-3937, doi:10.1128/AEM.69.7.3932-3937.2003.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, *et al.* (2016). Vegan: community ecology package.

Oliver PG, Holmes AM (2006). A new species of *Lucinoma* (Bivalvia: Lucinoidea) from the oxygen minimum zone of the Oman Margin, Arabian Sea. J Conchol 39:63-77.

Oliveros JC (2007). Venny. an interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/.

Espinosa E, Tanguy A, Le Panse S, Lallier F, Allam B, Boutet I (2013). Endosymbiotic bacteria in the bivalve *Loripes lacteus*: localization, characterization and aspects of symbiont regulation. J Exp Mar Biol Ecol 448:327-336, doi:10.1016/j.jembe.2013.07.015.

Pannebakker BA, Loppin B, Elemans CP, Humblot L, Vavre F (2007). Parasitic inhibition of cell death facilitates symbiosis. Proc Natl Acad Sci U S A 104:213-215, doi:10.1073/pnas.0607845104.

Papke RT, Gogarten JP (2012). How bacterial lineages emerge. Science 336:45-46, doi:10.1126/science.1219241.

Park S, Pham VH, Jung M, Kim S, Kim J, Roh D*, et al.* (2011). *Thioalbus denitrificans* gen. nov., sp. nov., a chemolithoautotrophic sulfur-oxidizing gammaproteobacterium, isolated from marine sediment. Int J Syst Evol Microbiol 61:2045-2051, doi:10.1099/ijs.0.024844-0.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043-1055, doi:10.1101/gr.186072.114.

Peek AS, Vrijenhoek RC, Gaut BS (1998). Accelerated evolutionary rate in sulfur-oxidizing endosymbiotic bacteria associated with the mode of symbiont transmission. Mol Biol Evol 15:1514-1523.

Peng Y, Leung HC, Yiu SM, Chin FY (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420-1428, doi:10.1093/bioinformatics/bts174.

Perez M, Juniper SK (2016). Insights into symbiont population structure among three vestimentiferan tubeworm host species at eastern Pacific spreading centers. Appl Environ Microbiol 82:5197-5205, doi:10.1128/AEM.00953-16.

Peters, JW, Schut, GJ, Boyd, ES, Mulder, DW, Shepard, EM, Broderick, JB, King, PW, Adams, MWW (2015). [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. BBA 1853: 1350-1369, doi: 10.1016/j.bbamcr.2014.11.021.

Petersen JM, Kemper A, Gruber-Vodicka H, Cardini U, van dG, Kleiner M*, et al.* (2016). Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. Nat Microbiol 2:16195, doi:10.1038/nmicrobiol.2016.195.

Petersen JM., Zielinski FU., Pape T, Seifert R, Moraru C, Amann R, *et al.* (2011). Hydrogen is an energy source for hydrothermal vent symbioses. Nature 476:176-80, doi:10.1038/nature10325.

Petrova OE, Schurr JR, Schurr MJ, Sauer K (2011). The novel *Pseudomonas aeruginosa* two-component regulator BfmR controls bacteriophage-mediated lysis and DNA release during biofilm development through PhdA. Mol Microbiol 81:767-783, doi:10.1111/j.1365-2958.2011.07733.x.

Pitcher RS, Watmough NJ (2004). The bacterial cytochrome cbb3 oxidases. Biochim Biophys Acta 1655:388-399, doi:10.1016/j.bbabio.2003.09.017.

Pomper BK, Saurel O, Milon A, Vorholt JA (2002). Generation of formate by the formyltransferase/hydrolase complex (Fhc) from *Methylobacterium extorquens* AM1. FEBS Lett 523:133-137, doi:S0014-5793(02)02962-9.

Ponnudurai R, Kleiner M, Sayavedra L, Petersen JM, Moche M, Otto A*, et al.* (2017). Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. ISME J 11:463-477, doi:10.1038/ismej.2016.124.

Poole RK, Hill S (1997). Respiratory protection of nitrogenase activity in *Azotobacter vinelandii* -- roles of the terminal oxidases. Biosci Rep 17:303-317.

Primavera JH, Lebata MJHL, Gustilo LF, Altamirano JP (2002). Collection of the clam *Anodontia edentula* in mangrove habitats in Panay and Guimaras, central Philippines. Wetlands Ecol Manage 10:363-370, doi:10.1023/A:1020983218203.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P*, et al.* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590-D596, doi:10.1093/nar/gks1219.

R Core Team (2016). R: a language and environment for statistical computing. https://www.r-project.org/.

Rader BA, Nyholm SV (2012). Host/microbe interactions revealed through "omics" in the symbiosis between the Hawaiian bobtail squid *Euprymna scolopes* and the bioluminescent bacterium *Vibrio fischeri*. Biol Bull 223:103-111, doi:10.1086/BBLv223n1p103.

Raina J, Eme L, Pollock FJ, Spang A, Archibald JM, Williams TA (2018). Symbiosis in the microbial world: from ecology to genome evolution. Biology Open 7:bio032524, doi:10.1242/bio.032524.

Read KRH (1965). The characterization of the hemoglobins of the bivalve mollusc *Phacoides pectinatus* (Gmelin). Comp Biochem Physiol 15:137-157.

Reid RGB, Brand DG (1986). Sulfide-oxidizing symbiosis in Lucinaceans: implications for bivalve evolution. Veliger 29:3-24.

Reynolds LK, Berg P, Zieman JC (2014). Lucinid clam influence on the biogeochemistry of the seagrass *Thalassia testudinum* sediments. Estuar Coast 30:482-490.

Rizzi M, Wittenberg JB, Coda A, Ascenzi P, Bolognesi M (1996). Structural bases for sulfide recognition in *Lucina pectinata* hemoglobin I. J Mol Biol 258:1-5, doi:10.1006/jmbi.1996.0228.

Robidart JC, Roque A, Song P, Girguis PR (2011). Linking hydrothermal geochemistry to organismal physiology: physiological versatility in *Riftia pachyptila* from sedimented and basalt-hosted vents. PLoS One 6:e21692, doi:10.1371/journal.pone.0021692.

Robinson MD, McCarthy DJ, Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140, doi:10.1093/bioinformatics/btp616.

Robinson MD, Oshlack A (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25, doi:10.1186/gb-2010-11-3-r25.

Roder C, Bayer T, Aranda M, Kruse M, Voolstra CR (2015). Microbiome structure of the fungid coral *Ctenactis echinata* aligns with environmental differences. Mol Ecol 24:3501-3511, doi:10.1111/mec.13251.

Rodrigues CF, Webster G, Cunha MR, Duperron S, Weightman AJ (2010). Chemosynthetic bacteria found in bivalve species from mud volcanoes of the Gulf of Cadiz. FEMS Microbiol Ecol 73:486-99, doi:10.1111/j.1574-6941.2010.00913.x.

Rodriguez-R LM, Konstantinidis KT (2014). Bypassing cultivation to identify bacterial species. Microbe 9:111-118.

Roeselers G, Newton IL (2012). On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. Appl Microbiol Biotechnol 94:1-10, doi:10.1007/s00253-011-3819-9.

Roeselers G, Newton IL, Woyke T, Auchtung TA, Dilly GF, Dutton RJ*, et al.* (2010). Complete genome sequence of *Candidatus* Ruthia magnifica. Stand Genomic Sci 3:163-173, doi:10.4056/sigs.1103048.

Rodriguez-R LM, Konstantinidis KT (2014). Bypassing cultivation to identify bacterial species. Microbe 9:111-118.

Rossi F, Colao E, Martinez MJ, Klein JC, Carcaillet F, Callier MD*, et al.* (2013). Spatial distribution and nutritional requirements of the endosymbiont-bearing bivalve *Loripes lacteus* (sensu Poli, 1791) in a Mediterranean *Nanozostera noltii* (Hornemann) meadow. J Exp Mar Biol Ecol 440:108-115, doi:10.1016/j.jembe.2012.12.010.

Roux S, Enault F, Hurwitz BL, Sullivan MB (2015). VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985, doi:10.7717/peerj.985.

Ruehland C, Blazejak A, Lott C, Loy A, Erseus C, Dubilier N (2008). Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. Environ Microbiol 10:3404-3416, doi:10.1111/j.1462-2920.2008.01728.x.

Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature 497:254-257, doi:10.1038/nature12048.

Sanders JG, Beinart RA, Stewart FJ, Delong EF, Girguis PR (2013). Metatranscriptomics reveal differences in *in situ* energy and nitrogen metabolism among hydrothermal vent snail symbionts. ISME J 7:1556-1567, doi:10.1038/ismej.2013.45.

Sanmartí N, Solé L, Romero J, Pérez M (2018). Seagrass-bivalve facilitative interactions: trait-mediated effects along an environmental gradient. Marine Environmental Research 133:99-104, doi:10.1016/j.marenvres.2017.12.002.

Schirmer T (2016). C-di-GMP synthesis: structural aspects of evolution, catalysis and regulation. J Mol Biol 428:3683-3701, doi:10.1016/j.jmb.2016.07.023.

Schleicher TR, Nyholm SV (2011). Characterizing the host and symbiont proteomes in the association between the bobtail squid, *Euprymna scolopes*, and the bacterium, *Vibrio fischeri*. PloS one 6:e25649, doi:10.1371/journal.pone.0025649.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB*, et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537-7541, doi:10.1128/AEM.01541-09.

Schreiber L, Kjeldsen KU, Funch P, Jensen J, Obst M, Lopez-Legentil S*, et al.* (2016). *Endozoicomonas* are specific, facultative symbionts of sea squirts. Front Microbiol 7:1042, doi:10.3389/fmicb.2016.01042.

Schweimanns M, Felbeck H (1985). Significance of the occurrence of chemoautotrophic bacterial endosymbionts in lucinid clams from Bermuda. Mar Ecol Prog Ser 24:113-120.

Scott KM, Schwedock J, Schrag DP, Cavanaugh CM (2004). Influence of form IA RuBisCO and environmental dissolved inorganic carbon on the delta13C of the clam-chemoautotroph symbiosis *Solemya velum*. Environ Microbiol 6:1210-1219, doi:10.1111/j.1462-2920.2004.00642.x.

Seah B (2014). Phylogenomics-tools. doi: 10.5281/zenodo.46122. https://github.com/kbseah/phylogenomics-tools/.

Shannon CE (1948). A mathematical theory of communication. Bell System Technical Journal 27:379-423, doi:10.1002/j.1538-7305.1948.tb01338.x.

Shapiro SS, Wilk MB (1965). An analysis of variance test for normality (complete samples). Biometrika 52:591-611, doi:10.1093/biomet/52.3-4.591.

Shuman KE, Hanson TE (2016). A sulfide:quinone oxidoreductase from *Chlorobaculum tepidum* displays unusual kinetic properties. FEMS Microbiol Lett 363:fnw100, doi:10.1093/femsle/fnw100.

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212, doi:10.1093/bioinformatics/btv351.

Simpson EH (1949). Measurement of diversity. Nature 163:688.

Small AL, McFall-Ngai MJ (1999). Halide peroxidase in tissues that interact with bacteria in the host squid *Euprymna scolopes*. J Cell Biochem 72:445-457, doi:10.1002/(SICI)1097-4644(19990315)72:43.0.CO;2-P.

Smejkalova H, Erb TJ, Fuchs G (2010). Methanol assimilation in *Methylobacterium extorquens* AM1: demonstration of all enzymes and their regulation. PLoS One 5:e13001, doi:10.1371/journal.pone.0013001.

Soto W, Nishiguchi MK (2014). Microbial experimental evolution as a novel research approach in the Vibrionaceae and squid-*Vibrio* symbiosis. Frontiers in Microbiology 5:593, doi:10.3389/fmicb.2014.00593.

Speare L, Cecere AG, Guckes KR, Smith S, Wollenberg MS, Mandel MJ*, et al.* (2018). Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. Proc Natl Acad Sci USA 115:E8528-E8537, doi:10.1073/pnas.1808302115.

Spiro B, Greenwood PB, Southward AJ, Dando PR (1986). $^{13}C/^{12}C$ ratios in marine invertebrates from reducing sediments: confirmation of nutritional importance of chemoautotrophic endosymbiotic bacteria. Mar Ecol Prog Ser 28:233-240.

Stamatakis A (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313, doi:10.1093/bioinformatics/btu033.

Stanley SM (2014). Evolutionary radiation of shallow-water Lucinidae (Bivalvia with endosymbionts) as a result of the rise of seagrasses and mangroves. Geology 42:803-806, doi:10.1130/G35942.1.

Stephens EA, Braissant O, Visscher PT (2008). Spirochetes and salt marsh microbial mat geochemistry: implications for the fossil record. Carnets de Géologie / Notebooks on Geology Article 2008/09:CG2008_A09.

Stewart FJ, Newton ILG, Cavanaugh CM (2005). Chemosynthetic endosymbioses: adaptations to oxic–anoxic interfaces. Trends Microbiol 13:439-448, doi:10.1016/j.tim.2005.07.007.

Stewart FJ, Cavanaugh CM (2006). Bacterial endosymbioses in *Solemya* (Mollusca: Bivalvia) -- model systems for studies of symbiont-host adaptation. Antonie Van Leeuwenhoek 90:343-360, doi:10.1007/s10482-006-9086-6.

Stewart FJ., Young CR, Cavanaugh CM (2008). Lateral symbiont acquisition in a maternally transmitted chemosynthetic clam endosymbiosis. Mol Biol Evol 25:673-687, doi:10.1093/molbev/msn010.

Stewart FJ, Dmytrenko O, Delong EF, Cavanaugh CM (2011). Metatranscriptomic analysis of sulfur oxidation genes in the endosymbiont of *Solemya Velum*. Frontiers in microbiology 2:134, doi:10.3389/fmicb.2011.00134.

Sun Y, O'Riordan MX (2010). Branched-chain fatty acids promote *Listeria monocytogenes* intracellular infection and virulence. Infect Immun 78:4667-4673, doi:10.1128/IAI.00546-10.

Sun YY, Chi H, Sun L (2016). *Pseudomonas fluorescens* filamentous hemagglutinin, an iron-regulated protein, is an important virulence factor that modulates bacterial pathogenicity. Front Microbiol 7:1320, doi:10.3389/fmicb.2016.01320.

Suomi T, Seyednasrollah F, Jaakkola MK, Faux T, Elo LL (2017). ROTS: An R package for reproducibility-optimized statistical testing. PLoS Comput Biol 13:e1005562, doi:10.1371/journal.pcbi.1005562.

Surger MJ, Angelov A, Stier P, Ubelacker M, Liebl W (2018). Impact of branched-chain amino acid catabolism on fatty acid and alkene biosynthesis in *Micrococcus luteus*. Front Microbiol 9:374, doi:10.3389/fmicb.2018.00374.

Suzuki MT, Taylor LT, DeLong EF (2000). Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5' nuclease assays. Appl Environ Microbiol 66:4605-4614, doi:10.1128/AEM.66.11.4605-4614.2000.

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS (2008). Distinct form I, II, III, and IV RuBisCO proteins from the three kingdoms of life provide clues about RuBisCO evolution and structure/function relationships. J Exp Bot 59:1515-1524, doi:10.1093/jxb/erm361.

Takamatsu N, Shiba T, Muramoto K, Kamiya H (1995). Molecular cloning of the defense factor in the albumen gland of the sea hare *Aplysia kurodai*. FEBS Lett 377:373-376, doi:10.1016/0014-5793(95)01375-X.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-2729, doi:10.1093/molbev/mst197.

Taylor JD, Glover EA (2000). Functional anatomy, chemosymbiosis and evolution of the Lucinidae. In: The Evolutionary Biology of the Bivalvia. Geological Society of London: London, pp 207-227.

Taylor JD, Glover EA (2009). A giant lucinid bivalve from the eocene of Jamaica - systematics, life habits and chemosymbiosis (Mollusca: Bivalvia: Lucinidae). Palaeontology 52:95-109, doi:10.1111/j.1475-4983.2008.00839.x.

Taylor JD, Glover EA (2010). Chemosymbiotic bivalves. In: The Vent and Seep Biota: Aspects from Microbes to Ecosystems. Springer: Netherlands, pp 107-128.

Taylor JD, Glover EA, Smith L, Dyal P, Williams ST (2011). Molecular phylogeny and classification of the chemosymbiotic bivalve family Lucinidae (Mollusca: Bivalvia). Zool J Linn Soc 163:15-49, doi:10.1111/j.1096-3642.2011.00700.x.

Taylor JD., Glover EA. (2013). New lucinid bivalves from shallow and deeper water of the Indian and West Pacific Oceans (Mollusca, Bivalvia, Lucinidae). ZooKeys 326:69-90, doi:10.3897/zookeys.326.5786.

Taylor J, Glover E, Smith L, Ikebe C, Williams S (2016). New molecular phylogeny of Lucinidae: increased taxon base with focus on tropical Western Atlantic species (Mollusca: Bivalvia). Zootaxa 4196:381-398, doi:10.11646/zootaxa.4196.3.2.

Taylor JD, Glover EA (2018). Hanging on — lucinid bivalve survivors from the Paleocene and Eocene in the western Indian Ocean (Bivalvia: Lucinidae). Zoosystema 40:123-142, doi:10.5252/zoosystema2018v40a7.

The UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res 43:D204-D212, doi:10.1093/nar/gku989.

Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.

Toft C, Fares MA (2008). The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. Mol Biol Evol 25:2069-2076, doi:10.1093/molbev/msn153.

Tomich M, Planet PJ, Figurski DH (2007). The *tad* locus: postcards from the widespread colonization island. Nat Rev Microbiol 5:363-375, doi:10.1038/nrmicro1636.

Toshchakov SV, Korzhenkov AA, Chernikova TN, Ferrer M, Golyshina OV, Yakimov MM*, et al.* (2017). The genome analysis of *Oleiphilus messinensis* ME102 (DSM 13489T)

reveals backgrounds of its obligate alkane-devouring marine lifestyle. Marine Genomics 36:41-47, doi:10.1016/j.margen.2017.07.005.

Turner S, Pryer KM, Miao VP, Palmer JD (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J Eukaryot Microbiol 46:327-338, doi:10.1111/j.1550-7408.1999.tb04612.x.

Udvardi M, Poole PS (2013). Transport and metabolism in legume-rhizobia symbioses. Annu. Rev. Plant Biol. 64:781–805. doi: 10.1146/annurev-arplant-050312-120235.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, *et al.* (2012). Primer3 - new capabilities and interfaces. Nucleic Acids Res 40:e115, doi:10.1093/nar/gks596.

van de Water JA, Melkonian R, Junca H, Voolstra CR, Reynaud S, Allemand D*, et al.* (2016). Spirochaetes dominate the microbial community associated with the red coral *Corallium rubrum* on a broad geographic scale. Sci Rep 6:27277, doi:10.1038/srep27277.

van der Geest M, Sall AA, Ely SO, Nauta RW, van Gils JA, Piersma T (2014). Nutritional and reproductive strategies in a chemosymbiotic bivalve living in a tropical intertidal seagrass bed. Mar Ecol Prog Ser 501:113-126, doi:10.3354/meps10702.

van der Heide T, Govers LL, de Fouw J, Olff H, van der Geest M, van Katwijk MM*, et al.* (2012). A three-stage symbiosis forms the foundation of seagrass ecosystems. Science 336:1432-1434, doi:10.1126/science.1219973.

Van Vranken JG, Na U, Winge DR, Rutter J (2015). Protein-mediated assembly of succinate dehydrogenase and its cofactors. Crit Rev Biochem Mol Biol 50:168-180, doi:10.3109/10409238.2014.990556.

Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, *et al.* (2015). Microbial species delineation using whole genome sequences. Nucleic Acids Res 43:6761-6771, doi:10.1093/nar/gkv657.

Veesenmeyer JL, Andersen AW, Lu X, Hussa EA, Murfin KE, Chaston JM*, et al.* (2014). NilD CRISPR RNA contributes to *Xenorhabdus nematophila* colonization of symbiotic host nematodes. Mol Microbiol 93:1026-1042, doi:10.1111/mmi.12715.

Vorholt JA (2002). Cofactor-dependent pathways of formaldehyde oxidation in methylotrophic bacteria. Arch Microbiol 178:239-249, doi:10.1007/s00203-002-0450-2.

Wallner G, Amann R, Beisker W (1993). Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. Cytometry 14:136-143, doi:10.1002/cyto.990140205.

Weis VM, Small AL, McFall-Ngai MJ (1996). A peroxidase related to the mammalian antimicrobial protein myeloperoxidase in the *Euprymna-Vibrio* mutualism. Proc Natl Acad Sci U S A 93:13683-13688.

Wilcoxon F (1945). Individual comparisons by ranking methods. Biometrics Bulletin 1:80-83, doi:10.2307/3001968.

Williams ST, Taylor JD, Glover EA (2004). Molecular phylogeny of the Lucinoidea (Bivalvia): non-monophyly and separate acquisition of bacterial chemosymbiosis. J Molluscan Stud 70:187-202, doi:10.1093/mollus/70.2.187.

Wittenberg JB, Wittenberg BA (1990). Mechanisms of cytoplasmic hemoglobin and myoglobin function. Annu Rev Biophys Biophys Chem 19:217-241, doi:10.1146/annurev.bb.19.060190.001245.

Wooldridge SA (2010). Is the coral-algae symbiosis really "mutually beneficial" for the partners? Bioessays 32:615–625, doi: 10.1002/bies.200900182.

Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO*, et al.* (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. Nature 443:950-955, doi:10.1038/nature05192.

Won YJ, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC (2003). Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. Appl Environ Microbiol 69:6785-6792.

Wu M, Eisen JA (2008). A simple, fast, and accurate method of phylogenomic inference. Genome Biol 9:R151, doi:10.1186/gb-2008-9-10-r151.

Xi H, Schneider BL, Reitzer L (2000). Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. J Bacteriol 182:5332-5341.

Yoshida N, Iguchi H, Yurimoto H, Murakami A, Sakai Y (2014). Aquatic plant surface as a niche for methanotrophs. Front Microbiol 5:30, doi:10.3389/fmicb.2014.00030.

Zaar A, Fuchs G, Golecki JR, Overmann J (2003). A new purple sulfur bacterium isolated from a littoral microbial mat, *Thiorhodococcus drewsii* sp. nov. Arch Microbiol 179:174-183, doi:10.1007/s00203-002-0514-3.

Zielinski FU, Pernthaler A, Duperron S, Raggi L, Giere O, Borowski C*, et al.* (2009). Widespread occurrence of an intranuclear bacterial parasite in vent and seep *Bathymodiolin* mussels. Environ Microbiol 11:1150-1167, doi:10.1111/j.1462-2920.2008.01847.x.