The Independent Partnership of *Thyasira* cf. *gouldi* and its Sulfur Oxidizing Symbiont

by Bonita McCuaig

A dissertation submitted to the School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in the Department of Biology

Memorial University of Newfoundland

December 2018

St John's, Newfoundland and Labrador, Canada

Abstract

Symbiotic relationships between invertebrates and chemoautotrophic bacteria can be found in many marine environments. Here we focus on describing the genomic content of the extracellular symbionts of *Thyasira* cf. *gouldi*. Previous work described three symbiont phylotypes designated A, B, and C. These bacteria are all closely related sulfur oxidising gammaproteobacteria.

Investigation of the ribulose-1,5-bisphosphate carboxylase (RuBisCo) gene showed a high level of diversity between bacterial populations hosted by *Thyasira* cf. *gouldi*. The RuBisCo had a higher diversity than the 16S rRNA gene and showed possible horizontal gene transfer events within phylotype A. A small number of host individuals also appeared to contain mixed populations of RuBisCo phylotypes not identified in 16S rRNA sequences.

Draft genomes were created from sequencing of host gill tissue. These samples were selected to include bacteria of each phylotype. The draft genomes created were highly fragmented, but many genes were identified. The metabolic capabilities were very similar and encode for carbon fixation through the Calvin-Benson-Basham cycle. Energy for this process is obtained through oxidation of sulfur compounds through both a reverse dissimilar sulfate reduction cycle and a bacterial sulfur oxidation cycle. The bacteria may also be able to utilize nitrogen or hydrogen oxidation for respiration. A functional tricarboxylic acid cycle was also identified in all three genomes suggesting the bacteria may be capable of heterotrophy. All metagenomes contained chemotaxis and flagellar genes.

An extracellular chromosome encoding a type IV secretion system was identified in the draft genome of phylotype B. If this plasmid is truly from the symbiont it could be used to export compounds to the host. Both phylotypes A and B contained genes involved in biofilm formation. A biofilm would benefit extracellular bacteria by creating a barrier from the host immune system and creating a controlled environment where the symbiont can exclude other environmental bacteria. Some genes associated with host interactions were also identified, although these were less conserved than the metabolic functions.

Our results support the evidence that *Thyasira* cf. *gouldi* obtains its symbionts from the environment. The draft genomes contained evidence of a flexible metabolism that can adapt to changing sediment conditions.

Acknowledgements

I would like to thank my co-supervisors Suzanne Dufour and Lourdes Peña-Castillo for all the support throughout this process, without their guidance I would not have been able to complete this research. I also want to thank Andrew Lang and his lab, for the use of equipment and training. In this regard Yunyun Fu was irreplaceable, thank you for all your help with sequencing. The Bonne Bay Marine Station made sampling possible, and I would like to thank the staff and students who helped with sampling. I would also like to thank Joost Verhoeven and Marta Canuti who were extremely helpful with data analysis.

I want to thank my family (Bonnie, Neil, and Nicolas McCuaig) for supporting my journey to Newfoundland. Moving so far away was hard but you guys made it easier. I also want to thank Corey Hutchings for being there when things got difficult. I would not have finished this dissertation without you guys.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

16S A	16S rRNA phylotype A
16S B	16S rRNA phylotype B
16S C	16S rRNA phylotype C
16S rRNA	a component of the 30S small subunit of the prokaryotic ribosome
ATP	adenosine triphosphate
Conitg	contiguous sequence
DNA	deoxyribonucleic acid
GC content	guanine and cytosine content
HGT	horizontal gene transfer
IUPAC	International Union of Pure and Applied Chemistry
OTU	operational taxonomic unit
PCR	polymerase chain reaction
RB 1	ribulose-1,5-bisphosphate carboxylase phylotype 1
RB 2	ribulose-1,5-bisphosphate carboxylase phylotype 2
RB 3	ribulose-1,5-bisphosphate carboxylase phylotype 3
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
RuBisCo	ribulose-1,5-bisphosphate carboxylase
Redox	reduction and/or oxidation

List of Appendices

Co-authorship Statement

I am the primary author of all chapters. I collected laboratory data (with the exception specified below) and performed all analyses. In all cases, the original manuscripts were written by me with the editorial input of my co-authors.

Chapter 2: I performed the16S rRNA PCR, and France Liboiron (Honours student) conducted the PCR for RuBisCo, under my supervision. I performed sequence analysis, wrote the chapter, and created figures, with editorial input from Suzanne Dufour and Lourdes Peña-Castillo. These results were published in PeerJ[1], and suggestions from the reviewers were incorporated into this chapter.

Chapter 3: I performed genomic sequencing, assembly, and annotation with guidance from Yunyun Fu and Lourdes Peña-Castillo. I wrote the chapter, with editorial input from Suzanne Dufour and Lourdes Peña-Castillo. This chapter was submitted to PlosONE and suggestions from the reviewers and editor were incorporated in this thesis chapter.

Chapter 4: I conducted genomic sequencing, assembly, annotation, and comparison with guidance from Yunyun Fu and Lourdes Peña-Castillo. The manuscript was edited by Suzanne Dufour, Lourdes Peña-Castillo, and Andrew Lang.

[1] McCuaig B, Liboiron F, Dufour SC. 2017. The bivalve *Thyasira* cf. *gouldi* hosts chemoautotrophic symbiont populations with strain level diversity. *PeerJ* 5:e3597.

Chapter 1: Introduction and Overview

1.1 Symbiosis

The term symbiosis is used to describe any relationship where distinct species live in close contact with each other for large portions of their life cycle. This definition includes mutualistic, commensal, and parasitic relationships, although symbiosis is commonly associated with mutualistic relationships (Sachs et al., 2011). Mutualistic relationships have been described between bacteria and animals from all major phyla (Douglas, 2014), and close contact has caused changes to both bacteria and their host over evolutionary time. Symbiotic relationships with prokaryotes are ubiquitous in nature and played major roles in shaping the evolution of eukaryotes, both the mitochondria and chloroplast are ancient endosymbionts that became cellular organelles (Margulis, 1981; Gil et al., 2010).

While every animal is surrounded by bacteria, most of these relationships are transient and do not persist across host generations. A symbiosis that becomes heritable, with symbionts passed from generation to generation, can lead to new selective pressures on both organisms (McFall-Ngai et al., 2013; Bennett and Moran, 2015), and can be a driving force for evolution (Brucker and Bordenstein, 2012). Stable symbiotic relationships can open new niches for hosts, allowing them to live off diets or move into habitats that could not sustain their non-symbiotic counterparts; this can cause effective ecological isolation and lead to speciation over time (Brucker and Bordenstein, 2012). To understand how symbioses affect the organisms involved, it is important to study relationships at various evolutionary stages, and not only focus on the well-established and highly derived obligate symbioses.

1.1.1 Evolution of symbiosis

Many of the original theories describing symbiont evolution focused on parasitic interactions because of their impact on human health and economics (Van Valen, 1974). Theories regarding parasitic relationships viewed them as an evolutionary "arms race". Parasites had to evolve quickly to avoid being destroyed by the host's immune system. The host would evolve new immune responses to counteract the parasites evolution, and in this way the two organisms would evolve quickly; each trying to out maneuver the other. This arms race was first described in the Red Queen Hypothesis (Van Valen, 1974). Theories explaining the evolution of mutualistic relationships were developed later and built upon the parasitic models (e.g. the Black Queen Hypothesis, which was derived from the Red Queen Hypothesis; (Morris et al., 2012). Early researchers viewed beneficial relationships as being the reverse of parasitic relationships, leading to the "mutualistic environment" theory (Law and Lewis, 1983). A mutualistic environment was thought to select for genomic stasis in the partners by positive frequency-dependant selection: the host and mutualistic symbiont evolve to accommodate the most common genotype in their partner with little genomic change over time (Law and Lewis, 1983; Sachs et al., 2011). These selective forces were thought to favor asexual reproduction and low recombination rates in the bacteria but not in the host, which requires sexual reproduction to adapt to shifting environmental conditions. It was believed that shifting conditions only affect the host because the bacteria are within an unchanging environment inside the host (Law and Lewis, 1983; Sachs et al., 2011). However, an increase in the amount of genomic data available to researchers has brought these historical models under new scrutiny (Ochman and Moran, 2001; Sachs et al., 2011). Recent whole genome studies have shown that both parasitic and mutualistic bacteria undergo rapid molecular changes and share a range of genomic features. They often use the same molecular mechanisms for invasion/colonization of their host and genomic reduction is

common to both obligate parasites and symbionts (Ochman and Moran, 2001; Moran and Plague, 2004; Wernegreen, 2005). These unexpected similarities challenge the idea of mutualistic and parasitic relationships being opposites and have led to the understanding that even in purely mutualistic relationships, the "partners" are not working in harmony. Each organism "wins" by getting the most out of the other while giving back as little as possible (Ferriere et al., 2002; Sachs et al., 2011; Douglas and Werren, 2016). It is clear that evolution within symbiotic relationships is much more complicated than originally thought.

1.2 Invertebrate-Bacteria Nutritional Symbiosis

Invertebrates from a variety of environments rely on symbiotic bacteria for nutritional needs. In these relationships the nutrients the bacteria provide can vary from specific amino acids or vitamins that are lacking in the host's diet, such as in phloem feeding insects or blood sucking parasites, to providing the host with most of their nutrients, as in some marine chemoautotrophic symbioses (Cavanaugh et al., 2006; Gündüz and Douglas, 2009). These relationships can open new environmental niches for the host and allow them to thrive on nutrient poor diets, removing competition from individuals without the symbiont (Brucker and Bordenstein, 2012; Joy, 2013). Aphids that specialize in feeding on phloem fluid sustain themselves on a diet low in vitamins and rely on complex relationships with symbiotic bacteria to supplement their diet. Aphid symbionts have severely reduced genomes yet maintain pathways to synthesize all 10 essential amino acids and many vitamins; this has allowed the insect to be sustained by phloem sap instead of the more nutritious xylem sap (Gündüz and Douglas, 2009). Nutritional symbioses can lead to radiation events, where many species arise quickly as the host moves into the new ecological niches (Brucker and Bordenstein, 2012).

Symbionts may evolve to supply a very small part of the host's nutrition, for instance, they may provide only the few amino acids that are missing in the host diet (Hansen and Moran,

2011). This shows that natural selection can act on the symbiont genome, leading to a perfect specialization for the niche of supplementing the host's diet. In other nutritional symbioses, the host receives a wide range of nutrients from the symbiont. In these cases, nutrient transfer is often achieved through direct digestion characterized by phagocytosis of the bacteria, followed by digestion within lysozyme vacuoles. This has been observed in some species of thyasirid clams, lucinid clams, bathymodioline mussels and other bivalves (Le Pennec et al., 1988; Espinosa et al., 2013; Laurich et al., 2017).

When a wide variety of nutrients are received from the symbiont, the host can rely on them for various amounts of their nutritional needs. Some hosts are also able to obtain food from the environment. In these facultative relationships, the symbiont may act more as a food reserve, supplementing the host's diet when environmental nutrients are scarce (Dufour and Felbeck, 2006). Complete reliance on symbionts for nutrients can lead to morphological changes within the host, such as the reduction or loss of the digestive tract (Reid and Bernard, 1980; Felbeck, 1981). Once this occurs the symbiosis has become obligate for the host. This is often associated with vertical transmission, (which is discussed in a later section); however, this is not always the case. Adult *Riftia pachytila* have no digestive system and completely relies on its symbionts for nutrition but acquires these symbionts from the environment rather than through vertical transmission (Nussbaumer et al., 2006). This appears to be a risky strategy: if a juvenile worm is not infected by an appropriate symbiont, it will perish.

1.3 Chemosymbiosis

Chemosymbiosis is common in marine environments where reduced and oxidizing conditions meet. The first chemosymbioses were identified in giant tubeworms from deep-sea hydrothermal vents in 1981 (Felbeck, 1981). At that time, the recent discovery of large numbers of macrofauna around deep-sea vents was perplexing: without substantial photosynthetic energy

inputs how could these organisms be so abundant and grow to such a large size? The giant tubeworm *Riftia pachyptila* was found to lack a mouth and digestive system yet dominated the ecosystem. It was in 1981 that Colleen Cavanaugh identified dense populations of bacteria within *Riftia pachyptila* body tissues (Cavanaugh et al., 1981). Horst Felbeck later showed that these bacteria were chemoautotrophic, oxidizing reduced sulfur compounds for energy and supplying their host with all the nutrients necessary for their surprising growth (Felbeck, 1981). Since this discovery, similar chemosymbioses have been found at cold seeps, whale falls, and mud volcanoes as well as in other marine environments with high redox potentials, such as anaerobic sediments and sea grass beds (Goffredi et al., 2003; Duperron et al., 2007; Gros et al., 2012; Batstone and Dufour, 2016). Symbioses from more accessible areas are now being studied; species from shallower areas are much easier to keep alive in the lab and have been used in controlled tank-based experiments (Ohishi et al., 2016; Zanzerl and Dufour, 2017).

Bivalves from many habitats have been shown to host symbiotic bacteria, usually on or within the gills which show some modifications (Dufour, 2005; Batstone et al., 2014). In bivalves that host their bacteria intracellularly, modified cells called bacteriocytes are found at the abfrontal end of the gill filaments (Distel, 1998). Bacteriocytes are modified cells that contain large vacuoles that hold the symbiotic bacteria. In less obligate symbioses, the bacteria are not intracellular, but are held between extracellular pockets limited by microvilli of gill epithelial cells (that are typically also called bacteriocytes) (Dufour, 2005). The gills in species with extracellular symbionts are often enlarged, with an expanded bacteriocyte zone increasing the surface area for hosting bacteria (Dufour, 2005; Batstone et al., 2014). Among bivalves, the interdependency of host-symbiont relationships varies greatly. Some associations are obligate for both parties (they are not found separately in nature), whereas in others, the bacterial symbionts

can have a free-living existence. In hosts that get most of their nutrition from their symbionts the digestive system is often reduced, and in extreme cases it is lost entirely (e.g. in *Solemya reidi*) (Stewart and Cavanaugh, 2006; Roeselers and Newton, 2012).

In the most evolved symbiotic relationships, the bacteria are passed vertically from parent to offspring through the gametes, ensuring transfer across generations. Some of the best studied bivalve symbioses fall within this category, and the relationship is obligate to both partners. Some clams within the genus *Solemya* are not found without their symbiotic bacteria and rely on them for all their nutritional needs (Stewart and Cavanaugh, 2006). In these relationships, because the bacteria cannot live outside the host, the symbionts have been isolated from free-living relatives and closely related bacteria hosted by closely related clams, causing co-speciation. (Moran et al., 2008).

1.4 Modes of Transmission

Three main modes of symbiont transmission have been described: 1) vertical transmission, where symbionts are passed directly from parent to offspring and are never exposed to the environment; 2) horizontal transmission, where the young are born without symbionts and are exposed to the symbionts from adult populations; and 3) environmental transmission, where the young are inoculated with symbionts from a free-living bacterial population (Bright and Bulgheresi, 2010; Sachs et al., 2011). While all these modes of transmission have been documented in nature, it is important to note that symbioses may incorporate more than one of these modes of symbiont transmission.

1.4.1 Vertical Transmission

In many of the relationships where the bacteria have become essential to the host, vertical transmission of symbionts within host gametes occurs, most commonly within the female gametes (Bright and Bulgheresi, 2010). Only a small number of symbionts can be transferred in

this manner, reducing the effective population size during transfer to subsequent host generations. A small population size allows genetic drift and founder effects to have a greater effect on genomic evolution (Moran, 2003). Vertical transmission leads to a number of genomic changes not usually seen in free-living populations; these changes are typically in the form of genomic reduction, which has been well documented in many symbionts that are unable to live outside their host (Moran and Plague, 2004; Moran et al., 2008; Bennett and Moran, 2015).

Genomic reduction occurs because selective pressures within the host are vastly different than the selective pressures that acted on the bacteria outside the host. Many of the genes required for a free-living lifestyle are no longer needed once the bacteria take up permanent residence within a dedicated host, and without selective pressures to maintain these genes, they are often lost (Moran and Plague, 2004). It has been proposed that there is a deletion bias within all genomes, such that it is much more common for DNA to be lost than gained (Ochman and Moran, 2001; Moran, 2003; Morris et al., 2012). Genomic bottlenecks also speed this process; as new generations may be inoculated with only a few bacteria, rare mutants can become common very quickly (Wernegreen and Moran, 1999).

Early in the process of genome reduction following host restriction, many genes lose their function and become pseudogenes. There is also a proliferation of mobile elements within the genome (Moran and Plague, 2004). Mobile elements can play a key role in generating pseudogenes, as well as playing a role in larger changes such as deletions and insertions. If two mobile elements of the same family are close in the genome, there may be an excision mistake, cutting at the far edges of the mobile elements and carrying along the genomic DNA in between. If this sequence fails to re-enter the genome, a large deletion occurs. The sequence can also move

within the genome, or be inverted (Wernegreen and Moran, 1999; Ochman and Moran, 2001; Moran, 2003; Moran and Plague, 2004).

After the initial proliferation of mobile elements and wide-spread loss of function in genes the genome begins to reduce itself in earnest. Without recombination to spread them the mobile elements are slowly lost. After a loss of function, there is no selective pressure on the pseudogenes and they are slowly degraded by the deletion bias (Moran et al., 2008). If mutations occur in DNA repair genes this can speed the process of genomic reduction, and a low GC (<40%) content within the genome. It was found that mutations of *mutY* were common in *Calyptogena* symbionts from multiple species. Those symbionts with non-functioning *mutY* genes had a much lower GC content, because this gene corrects mutations to adenine and thymine back to guanine and cytosine (Kuwahara et al., 2011). Obligate symbionts often have a low GC content. *Hodgkinia cicadicola* is a notable exception: they are obligate symbionts with a high GC content (58.4%), but reduced genomes (Kuwahara, et al., 2011). The inactivation of DNA repair genes increases the speed of gene mutation and increased genomic reduction.

1.4.2 Horizontal Transmission

Horizontal transmission requires that the host releases viable symbionts, which can be taken up by a juvenile of the same species. Symbiont release may occur when the host is alive, or symbionts may escape after the death of the host. True horizontal transmission is observed in insect symbioses, but is rare in marine bivalves, occurring mostly alongside vertical or environmental transmission (Brissac et al., 2009; Stewart et al., 2009). Bivalve horizontal transmission events have been shown in the generally vertically transmitted symB symbiont type of *Vesicomya* sp. (Stewart et al., 2009). Rare horizontal transmission events within a largely vertically transmitted population would help counteract the effects of genomic reduction and increase the genetic diversity of the symbiont population.

When coupled with environmental transmission, the horizontal events release live symbionts into the environment, where they join the existing free-living population. This can increase genetic diversity for the free-living bacterial population, particularly if the host has transported symbionts from a different habitat. Theoretically, the adults could also inoculate a new habitat with symbionts, ensuring some bacteria are available for juveniles entering the area.

1.4.3 Environmental Transmission

Environmental transmission occurs when juvenile hosts are inoculated with symbionts from a free-living population of bacteria. These symbiotic bacteria can flourish in the environment and do not require the host (Harmer et al., 2008; Dufour et al., 2014). These types of symbiotic relationships are often facultative, with the host retaining some ability to feed on its own. Notable exceptions are the deep-sea tubeworms *Riftia pachyptila* and *Tevnia jericonana* whose adult stages have no digestive system and rely entirely on environmentally transmitted symbionts (Nussbaumer et al., 2006). The bacteria in these relationships retain a diverse breeding population and show no signs of genomic reduction. They also retain genes for a free-living lifestyle and need to navigate in the external environment; therefore, genes for chemotaxis and flagella are often present (Robidart, 2006; Robidart et al., 2008; Dmytrenko et al., 2014). It is also common for facultative bacterial symbionts to have a diverse metabolism, able to deal with different redox conditions, fix carbon, and sometimes live heterotrophically as well (Dmytrenko et al., 2014; Reveillaud et al., 2018).

1.5 Thyasirid bivalves

Thyasirid clams can be found in oceans around the world at a large range of depths but are restricted to cold waters. They are often associated with oxygen minimum zones, such as cold seeps, hydrothermal vents, mud volcanoes, and organically enriched sediments (Rozemarijn, 2011). Although thyasirid bivalves have long been grouped with chemosymbiotic

lucinids in the order Lucinoidea, recent molecular studies have shown showed that families are not closely related, and the superfamily Thyasiroidea was erected (Taylor et al., 2007a). Phylogenetic studies place the Thyasiridae near the base of the heterodont bivalves (Taylor et al., 2007b; Combosch et al., 2017). The fossil record agrees with this placement, with the oldest thyasirid fossils being found in the early Cretaceous. Many early fossils are found at putative cold seep and hydrothermal vent sites (Kiel, 2008; Hryniewicz et al., 2017). It is likely that the thyasirids evolved earlier (Taylor et al., 2007a) and possibly outside reducing environments; however, due to their small size and indistinct shell morphology, earlier thyasirid fossils have not been reported.

Thyasiroidea contains 11 of genera, including both symbiotic (genera with symbionts) and asymbiotic (genera without symbionts) groups (Rozemarijn et al., 2011). Thyasirids are morphologically diverse; among bivalves, the Thyasiridae is the only family in which a variable number of gill demibranchs can be found. Dufour (2005) identified three distinct gill types within the family, and with an additional "tubular" morphology being described more recently (Oliver, 2014). Further, the genus *Thyasira* is unique among bivalves in being the only one known to contain both symbiotic and asymbiotic species. Although the high level of diversity led to some investigators questioning the phylogenetic validity of *Thyasira* molecular evidence has supported the placement of species within the genera (Taylor et al., 2007a).

Thyasirid shells are plain and lack many defining features, notably dentition. They do, however, show some distinguishing characters, including the presence of a prominent posterior sulcus in many species. The internal structure shows more distinct morphological features, including the visceral pouches that enclose the digestive gland and tubules, and, in many species, an elongate foot (Oliver and Killeen, 2002). Thyasirids do not have siphons and use their

extensible foot to create a single inhalent/exhalent tube to maintain contact with overlying water. Many thyasirid species also use their foot to construct elaborate, ramifying burrows, which can be used to access sulfide pockets in the anoxic sediments (Dufour and Felbeck, 2003).

1.6 *Thyasira* cf. *gouldi*

Thyasira gouldi was first described off the coast of Massachutsetts in 1845 (Philippi, 1845). This species has been described in both the North Atlantic and the North Pacific oceans. Clams closely resembling *Thyasira gouldi* were identified in Bonne Bay, Newfoundland, Canada (Batstone *et al.* 2014). Upon closer investigation, Bonne Bay *Thyasira* cf. *gouldi* specimens were found to form a complex of 3 closely related operational taxonomic units (OTUs), with OTUs 1 and 2 being symbiotic, while no bacteria were found associated with OTU 3 (Batstone et al., 2014).

Behavioural investigations showed that both symbiotic and asymbiotic *Thyasira* cf. *gouldi* from Bonne Bay create complex burrows (called pedal tracts) in surrounding sediments. Pedal tracts were thought to be a means of "mining" the reduced sulfide compounds required by the symbionts (Zanzerl, 2015), and the presence of this behaviour in asymbiotic species was puzzling. It now appears that pedal tract formation, coupled with bioirrigation of the biogenic structures, promotes the growth of sulfur oxidizing bacteria along burrow walls, and that thyasirids may collect these bacteria on the mucus lining of their foot, bring them to the mantle cavity, and consume them (Zanzerl and Dufour, 2017). Pedal feeding explains the similarity in isotopic profiles of asymbiotic thyasirids, which may gain a large portion of their nutrients from chemoautotrophic bacteria (Zanzerl, 2015). Symbiotic thyasirids may also acquire their symbionts by collecting them from burrow linings using their foot; this could explain the presence of magnetosomes (structures that some free-living sulfur-oxidizing bacteria employ to assist the location of oxic-anoxic interfaces, such as those present around burrow linings) in *T.*

cf. *gouldi* symbionts (Dufour et al., 2014). In thyasirids, bacterial farming may have been the precursor to chemosymbiosis. Asymbiotic thyasirids can live in close contact with sulfur oxidizing bacteria and promote their proliferation through their irrigation behaviours. It is plausible that some of these bacteria, having been brought into contact with the bivalve's gills during pre-ingestive processing, established residence among the microvilli of gill epithelial cells (Zanzerl and Dufour, 2017).

The bacteria associated with symbiotic *Thyasira* cf. *gouldi* were identified as Gammaproteobacteria, closely related to the sulfur-oxidizing symbionts of *Thyasira flexuosa*. The 16S rRNA sequences of these two symbionts were significantly different, suggesting different bacterial species (Batstone and Dufour, 2016). Additionally, 16S rRNA sequencing of multiple *T.* cf. *gouldi* specimens revealed the presence of three distinct bacterial phylotypes (A, B, and C; Batstone and Dufour, 2016). There appeared to be no co-speciation between clam OTU and bacterial phylotype, and sequencing suggested the presence of a single bacterial phylotype per host clam. Additional PCR mediated sequencing showed that *T.* cf. *gouldi* symbionts contained a form II RuBisCo gene (Dufour et al., 2014). This enzyme is a key component of the Calvin Benson Bassham Cycle, which is used for carbon fixation. RuBisCo form II is less efficient than form I but is better suited to low oxygen levels (Badger and Bek, 2008). The symbionts of *Thyasira* cf. *gouldi* likely encounter low oxygen levels both in their sedimentary environment, and periodically in the host's gills (when hypoxic or anoxic water is pumped throughout burrow structures and into the mantle cavity; Hakonen et al., 2010).

1.7 Research Objectives

The main purpose of this study was to explore the genome of thyasirid symbionts, in order to: 1) describe features that might relate to the extracellular and facultative nature of these symbionts; 2) characterize potential metabolic capabilities in these symbionts; and 3) gain

insights on host-symbiont relationships and potential interactions. The work focuses on the gammaproteobacterial symbionts of *Thyasira* cf. *gouldi* from Bonne Bay, Newfoundland, and uses PCR amplification and sequencing, as well as next-generation shotgun sequencing, to explore the genome of these poorly-known symbionts.

In Chapter 2, we investigate the potential of hidden heterogeneity within the symbiont population of a single host. The 16S rRNA sequences is very stable and evolves more slowly than most other genes. Bacteria sharing a 16S rRNA sequence can have very different metabolic genes within their genomes (Ikuta et al., 2016). By investigating the sequence of a more quickly evolving gene (RuBisCo), we investigate the possibility of different gene phylotypes within the symbiont population of a single host clam.

Chapter 3 contains a preliminary description of the genomic features of symbiont phylotype B. This is the first attempt to create a draft genome of a thyasirid symbiont, and to the authors' knowledge the first description of an extracellular bivalve symbiont genome. The main focus of this chapter is to create a general description of the symbiont's metabolic capabilities and identify any genes that may be involved in host interactions and continuation of the symbiosis.

In Chapter 4 we compare the similarities and differences between the draft genomes created for phylotypes A, B, and C. By understanding which genes are conserved and which genes differ between the phylotypes, we can make some inferences about the symbiotic relationship. We also look to identify genes that are present in all three genomes that may be involved in host/symbiont identification, colonization, and the persistence of the symbiosis.

1.8 References

Badger MR, Bek EJ. 2008. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. *Journal of experimental botany* 59:1525-1541.

Batstone RT, Dufour SC. 2016. Closely related thyasirid bivalves associate with multiple symbiont phylotypes. *Marine Ecology* 37:988-997.

Batstone RT, Laurich JR, Salvo F, Dufour SC. 2014. Divergent chemosymbiosis-related characters in *Thyasira* cf. *gouldi* (Bivalvia: Thyasiridae). *PLoS ONE* 9:e92856

Bennett GM, Moran NA. 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proceedings of the National Academy of Sciences* 112:10169-10176.

Bright M, Bulgheresi S. 2010. A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology* 8:218-230.

Brissac T, Gros O, Merçot H. 2009. Lack of endosymbiont release by two Lucinidae (Bivalvia) of the genus Codakia: consequences for symbiotic relationships. *FEMS Microbiology Ecology* 67:261-267.

Brucker RM, Bordenstein SR. 2012. Speciation by symbiosis. *Trends in Ecology and Evolution* 27:443-451.

Cavanaugh C, McKiness ZP, Newton ILG, Stewart F. 2006. Marine Chemosynthetic Symbioses. In: *The Prokaryotes* (eds. by Dworkin M, Falkow SI, Rosenberg E, Schleifer K-H & Stackebrandt E), pp. 475-507. Springer, New York.

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*, 340-342.

Combosch DJ, Collins TM, Glover EA, Graf DL, Harper EM, Healy JM, Kawauchi GY, Lemer S, McIntyre E, Strong EE. 2017. A family-level tree of life for bivalves based on a Sanger-sequencing approach. *Molecular Phylogenetics and Evolution* 107:191-208.

Distel DL. 1998. Evolution of chemoautotrophic endosymbioses in bivalves. *Bioscience* 48:277-286.

Dmytrenko O, Russell S, Loo WT, Fontanez KM, Liao L, Roeselers G, Charma R, Stewart F, Newton ILG, Woyke T, Wu D, Lang JM, Eisen JA, Cavanaugh C. 2014. The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. *BMC Genomics* 15:1-20.

Douglas AE. 2014. Symbiosis as a general principle in eukaryotic evolution. *Cold Spring Harbor Perspectives in Biology* 6: a016113.

Douglas AE, Werren JH. 2016. Holes in the hologenome: Why host-microbe symbioses are not holobionts. *American Society for Microbiology* 7.

Dufour SC. 2005. Gill anatomy and the evolution of symbiosis in the bivalve family Thyasiridae. *Biological Bulletin* 208:200-212.

Dufour SC, Felbeck H. 2003. Sulphide mining by the superextensile foot of symbiotic thyasirid bivalves. *Letter to Nature* 426:65-67.

Dufour SC, Felbeck H. 2006, Symbiont abundance in thyasirids (Bivalvia) is related to particulate food and sulphide availability. *Marine Ecology Progress Series* 320:185-194.

Dufour SC, Laurich JR, Batstone RT, McCuaig B, Elliot A, Poduska KM. 2014. Magnetosome-containing bacteria living as symbionts of bivalves. *The ISME Journal* 8:2453-2462.

Duperron S, Fiala-Medioni A, Caprais J-C, 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 Microbiology Ecology* 59:64-70.

Espinosa EP, 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. *Journal of Experimental Marine Biology and Ecology* 448:327-336.

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

Ferriere R, Bronstein JL, Rinaldi S, Law R, Gauduchon M. 2002. Cheating and the evolutionary stability of mutualisms. *Proceedings of the Royal Society of London B: Biological Sciences* 269:773-780.

Gil R, Latorre A, Moya A. 2010. Evolution of prokaryote-animal symbiosis from a genomics persepective. In: *(Endo)symbiotic Methanogenic Archaea* (ed. by Hackstein JHP), pp. 208-233. Springer-Verlag, Berlin.

Goffredi S, Hurtado L, Hallam S, Vrijenhoek R. 2003, Evolutionary relationships of deep-sea vent and cold seep clams (Mollusca: Vesicomyidae) of the" pacifica/lepta" species complex. *Marine Biology* 142:311-320.

Gros O, Elisabeth NH, Gustave SDD, Caro A, Dubilier N. 2012. Plasticity of symbiont acquisition throughout the life cycle of the shallow-water tropical lucinid *Codakia orbiculata* (Mollusca: Bivalvia). *Environmental Microbiology* 14:1584-1595.

Gündüz EA, Douglas A. 2009. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proceedings of the Royal Society of London B: Biological Sciences* 276:987-991.

Hansen AK, Moran NA. 2011. Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of Sciences* 108:2849-2854.

Harmer TL, Rotjan RD, Nussbaumer AD, Bright M, Ng AW, DeChaine EG, Cavanaugh C. 2008. Free-living tube worm endosymbionts found at deep-sea vents. *Applied and Environmantal Microbiology* 74:3895-3898.

Hryniewicz K, Amano K, Jenkins R, Kiel S. 2017. Thyasirid bivalves from Cretaceous and Paleogene cold seeps. *Acta Palaeontologica Polonica* 62:705-728.

Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, Aoki Y, Inoue K, Teruya M, Satou K, Teruya K, Shimoji M, Tamotsu H, Hirano T, Maruyama T, Yoshida T. 2016. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. *ISME Journal* 10:990-1001.

Joy JB. 2013. Symbiosis catalyses niche expansion and diversification. *Proceedings of the Royal Society of London B: Biological Sciences* 280:20122820.

Kiel S. 2008. Fossil evidence for micro-and macrofaunal utilization of large nekton-falls: examples from early Cenozoic deep-water sediments in Washington State, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology* 267:161-174.

Kuwahara H, Takaki Y, Shimamura S, Yoshida T, Maeda T, Kunieda T, Maruyama T. 2011. Loss of genes for DNA recombination and repair in the reductive genome evolution of thioautotrophic symbionts of *Calyptogena* clams. *BMC Evolutionary Biology* 11.

Laurich JR, Dove R, Paillard C, Dufour SC. 2017. Life and death in facultative chemosymbioses: control of bacterial population dynamics in the Thyasiridae. *Symbiosis* 1-11.

Law R, Lewis D. 1983. Biotic environments and the maintenance of sex–some evidence from mutualistic symbioses. *Biological Journal of the Linnean Society* 20:249-276.
Le Pennec M, Diouris M, Herry A. 1988. Endocytosis and lysis of bacteria in gill epithelium of Bathymodiolus thermophilus, Thyasira flexuosa and Lucinella divaricata (Bivalve, Molluscs). *Journal of Shellfish Research* 7:483-489.
Margulis L. 1981. Symbiosis in cell evolution: Life and its environment on the early earth. Boston University, Boston, MA, United States.
McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences* 110:3229-3236.
Moran NM. 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. *Current Opinion in Microbiology* 6:512-518.
Moran NM, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* 42:165-190.
Moran NM, Plague GR. 2004. Genomic changes following host restriction in bacteria. *Current Opinions in Genetics and Development* 14:627-633.
Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *MBio* :e00036-12.
Nussbaumer AD, Fisher CR, Bright M. 2006. Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* 441:345-348.
Ochman H, Moran NM. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* 292:1096-1098.
Ohishi K, Yamamoto M, Tame A, Kusaka C, Nagai Y, Sugimura M, Inoue K, Uematsu K, Yoshida T, Ikuta T, Takashi T, Maruyama T. 2016. Long-term cultivation of the deep-sea clam *Calyptogena okutanii*: Changes in the abundance of chemoautotrophic symbiont, elemental sulfur, and mucus. *Biological Bulletin* 230:257-267.
Oliver G. 2014. "TUBULAR GILLS" Extreme gill modification in the Thyasiroidea with the description of *Ochetoctena tomasi* gen. et sp. nov.(Bivalvia: Thyasiroidea). *Zoosystematics and Evolution* 90:121.
Oliver PG, Killeen IJ. 2002. *The Thyasiridae (Mollusca: Bivalvia) of the British continental shelf and North Sea oil fields: an identification manual*. Wales: Studies in Marine Biodiversity and Systematics from the National Museum of p. 78. Wales.
Philippi R. 1845. Bemerkungen über die Mollusken-Fauna von Massachusetts. *Zeitschrift für Malakozoologie* 1845:68-79.
Reid RG, Bernard FR. 1980. Gutless bivalves. *Science* 208:609-610.
Reveillaud J, Anderson R, Reves-Sohn S, Cavanaugh C, Huber JA. 2018. Metagenomic investigation of vestimentiferan tubeworm endosymbionts from Mid-Cayman Rise reveals new insights into metabolism and diversity. *Microbiome* 6:19.
Robidart JC. 2006. Metagenomics of the Riftia pachyptila symbiont. In: *Marine Biology*, p. 128. UC San Diego.
Robidart JC, Bench S, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, Allen EE, Felbeck H. 2008. Metabolic versitilty of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environmental Microbiology* 10:727-737.
Roeselers G, Newton ILG. 2012. On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. *Applied and Environmantal Microbiology* 94:1-10.

Rozemarijn K, Christoffer S, Anders KJ, Endre W. 2011. Ecology of twelve species of Thyasiridae (Mollusca: Bivalvia). *Marine Pollution Bulletin* 62:786-791.
Sachs JL, Skophammer RG, Regus JU. 2011. Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences* 108:10800-10807.
Stewart F, Cavanaugh C. 2006. Bacterial endosymbiosis in *Solemya* (Mollusca: Bivalvia) - Model systems for studies of symbiont-host adaptation. *Antonie van Leeuwenhoek* 90:343-360.
Stewart F, Young CR, Cavanaugh C. 2009. Evidence for homologous recombination in intracellular chemosynthetic clam symbionts. *Molecular Biology and Evolution* 26:1391-1404.
Taylor JD, Williams ST, Glover EA. 2007a. Evolutionary relationships of the bivalve family Thyasiridae (Mollusca: Bivavlia), monphyly and superfamily status. *Journal of the Marine Biological Association of the United Kingdom* 87:565-574.
Taylor JD, Williams ST, Glover EA, Dyal P. 2007b. A molecular phylogeny of heterodont bivalves (Mollusca: Bivalvia: Heterodonta): new analyses of 18S and 28S rRNA genes. *Zoologica Scripta* 36:587-606.
Van Valen L. 1974. Molecular evolution as predicted by natural selection. *Journal of Molecular Evolution* 3:89-101.
Wernegreen J, Moran N. 1999. Evidence for genetic drift in endosymbionts (Buchnera): analyses of protein-coding genes. *Molecular Biology and Evolution* 16:83-97.
Wernegreen JJ. 2005. For better or worse: genomic consequences of intracellular mutualism and parasitism. *Current Opinion in Genetics and Development* 15:572-583.
Zanzerl H. 2015. The burrowing behaviour and diet of thyasirid bivalves from Bonne Bay, NL. MSc Thesis, *Department of Biology*, Memorial University of Newfoundland.
Zanzerl H, Dufour SC. 2017. The burrowing behaviour of symbiotic and asymbiotic thyasirid bivalves. *Journal of Conchology* 42:299.

Chapter 2: The bivalve *Thyasira* cf. *gouldi* hosts heterogeneous chemoautotrophic symbiont populations with strain level diversity[2]

Bonita McCuaig, France Liboiron, Suzanne C. Dufour

2.0 Abstract

Symbioses are common in nature. In some cases, marine invertebrates host bacteria capable of chemosynthesis, and often provide nutrients for their host. In chemosynthetic symbiosis both the mode of symbiont transmission and the site of bacterial housing can affect the composition of the symbiont population. Vertically transmitted symbionts, as well as those hosted intracellularly, are more likely to form clonal populations within their host. Conversely, symbiont populations that are environmentally acquired and extracellular may be more likely to be heterogeneous/mixed within host individuals, as observed in some mytilid bivalves. The symbionts of thyasirid bivalves are also extracellular, but limited 16S rRNA sequencing data suggest that thyasirid individuals contain uniform symbiont populations. In a recent study, *Thyasira* cf. *gouldi* individuals from Bonne Bay, Newfoundland, Canada were found to host one of three 16S rRNA phylotypes of sulfur-oxidizing gammaproteobacteria, suggesting environmental acquisition of symbionts and some degree of site-specificity. Here, we use Sanger sequencing of both 16S rRNA and the more variable ribulose-1,5-bisphosphate carboxylase (RuBisCo) PCR products to further examine *Thyasira* cf. *gouldi* symbiont diversity at the scale of host individuals, as well as to elucidate any temporal or spatial patterns in symbiont diversity within Bonne Bay, and any relationships with host OTU or size. We obtained symbiont 16S rRNA and RuBisCo form II sequences from 54 and 48 host individuals, respectively, during nine

[2] McCuaig B, Liboiron F, Dufour SC. 2017. The bivalve *Thyasira* cf. *gouldi* hosts chemoautotrophic symbiont populations with strain level diversity. *PeerJ* 5:e3597.

sampling trips to three locations over four years. Analyses uncovered the same three closely related 16S rRNA phylotypes obtained previously, as well as three divergent RuBisCo phylotypes; these were found in various pair combinations within host individuals, suggesting incidents of horizontal gene transfer during symbiont evolution. While we found no temporal patterns in phylotype distribution or relationships with host OTU or size, some spatial effects were noted, with some phylotypes being excluded from, or only found within, particular sampling sites. The sequencing also revealed symbiont populations within individual hosts that appeared to be a mixture of different phylotypes, based on multiple base callings at divergent sites. This work provides further evidence that *Thyasira* cf. *gouldi* acquires its symbionts from the environment and reveals that hosts can harbour symbiont populations consisting of multiple, closely related bacterial strains.

2.1 Introduction

Symbioses between animals and bacteria are ubiquitous and, in many cases, advantageous to the host (McFall-Ngai et al., 2013). Animals often benefit from symbiont-derived metabolic products, which may include essential or non-essential nutrients. In a class of animal-bacterial relationships called chemosymbioses, marine invertebrates receive nutrients from chemoautotrophic bacterial symbionts such as sulfur-oxidizing gammaproteobacteria. Since the discovery of chemosymbioses in giant tubeworms from hydrothermal vents (Felbeck, 1981; Jones, 1981), invertebrates from various phyla and marine habitats were found to establish symbioses with a wide diversity of chemoautotrophic bacteria (Cavanaugh et al., 2006; Dubilier et al., 2008). Chemosymbiotic host species differ in their degree of specificity for particular bacteria: some host species will associate with a single strain and harbour clonal populations, while others are more flexible and can form symbioses with more than one bacterial strain, in either single or mixed populations (Dubilier, et al., 2008).

The composition of a chemosymbiotic host's symbiont population is affected by the mode of symbiont transmission. Some symbionts are transmitted vertically between host generations, resulting in highly specific and co-evolved relationships (Dubilier et al., 2008). Vertically transmitted symbionts often form clonal populations within a host, as only a few bacteria are transmitted each generation, leading to population bottlenecks. Among vertically transmitted symbionts, gene flow is restricted, and recombination events are rare, further reducing the genetic variation within the population (Goffredi et al., 2003; Wernegreen, 2005; Caro et al., 2007; Dubilier et al., 2008); nevertheless, rare horizontal transmission events can occur (Stewart and Cavanaugh, 2009; Stewart et al., 2009). In environmentally transmitted symbioses, hosts are inoculated every generation from a free-living bacterial population (Nussbaumer et al., 2006; Won et al., 2008; Vrijenhoek, 2010). Hosts that acquire symbionts from their environment have the advantage of associating with a locally adapted strain and may harbour mixed populations of symbionts (Vrijenhoek et al., 2007; Dubilier et al., 2008; Moran et al., 2008; Ikuta et al., 2016). Symbiont populations may become increasingly specific over a host's lifetime: as lucinid bivalves mature, the most efficient strain of bacteria is selected from the original heterogeneous, environmentally acquired population (Brissac et al., 2016). The composition of the symbiont population in adult organisms is also affected by the amount of time a host is 'competent' and can be colonized by symbionts. Species such as *Riftia pachyptila*, *Anodontia alba*, and *Codakia pectinella* are only competent for a short time and are more likely to have clonal or nearly clonal populations as adults (Nussbaumer et al., 2006; Brissac et al., 2016). Conversely, some species are competent throughout their lifetime and continually acquire new strains of bacteria from the environment (Wentrup et al., 2014; Ikuta et al., 2016). Examining the degree of symbiont specificity in a host can inform us on their potential for

adaptation when faced with environmental change and help us understand how symbioses evolve and break down (Sachs et al., 2011).

Chemosymbiotic invertebrates differ in the degree to which symbionts are integrated in their bodies or cells, and this variation could affect specificity, especially in hosts that are colonized by free-living bacteria. In many invertebrates (e.g. vesicomyid bivalves and giant tubeworms; e.g. Cavanaugh et al., 2006) symbionts are maintained within host cells whereas in others, symbionts are internalized but extracellular (e.g. thyasirid and some bathymodiolin bivalves with symbionts held among microvilli of gill epithelial cells; e.g. Dufour, 2005; Duperron et al., 2008a), or are epibiotic, attached to the external surface of the body (e.g. nematodes; Ott et al., 1991). Among hosts with environmentally acquired symbionts, those that house symbionts intracellularly might be expected to show more specificity than those with external symbionts, as internalization processes could be selective (Brissac et al., 2016). Lucinid bivalves and vestimentiferan tubeworms are examples of hosts that are colonized by symbionts as juveniles, and contain one or two, metabolically divergent, locally sourced symbiont phylotype(s) as adults (Brissac et al., 2016). In contrast, hosts with extracellular symbionts might be expected to associate with a broader, more variable range of symbiont types; this has been observed in wood-fall mussels having mixed populations of 5-6 divergent phylotypes (Duperron et al., 2008b). A different situation has been observed more recently in thyasirid bivalves with extracellular symbionts: *Thyasira* cf. *gouldi* conspecifics from the same fjord (Bonne Bay, Newfoundland, Canada) associated with one of three highly similar 16S rRNA phylotypes of sulfur-oxidizing gammaproteobacteria (Batstone and Dufour, 2016). These observations highlight the fact that the mechanisms of symbiont selection in extracellular symbioses are not known and may differ markedly among host taxa.

Thyasira cf. *gouldi* inhabiting cold marine sediments within the glacial fjord of Bonne Bay (Canada) form a complex of three distinct operational taxonomic units (OTUs), with OTUs 1 and 2 having elongated gill filaments housing thioautotrophic bacteria, while OTU 3 has shorter gill filaments and is asymbiotic (Batstone et al., 2014). Symbiotic and asymbiotic OTUs of *Thyasira* cf. *gouldi* create elaborate burrows within the sediment using their extensible foot (Zanzerl and Dufour, 2017). In symbiotic thyasirids, burrow formation has been interpreted as a mechanism to "mine" for the sulfur compounds the symbiont requires (Dufour and Felbeck, 2003; Dando et al., 2004), while in some asymbiotic thyasirids, burrows have been associated with deposit feeding (Zanzerl and Dufour, 2017). Thyasirid irrigation leads to the establishment of oxic/anoxic interfaces around burrow linings (Dando et al., 2004; Hakonen et al., 2010) and likely favours colonization of sulfur-oxidizing bacteria. The presence of magnetosome particles in thyasirid symbionts suggests that, in their free-living state, symbionts navigate to burrow linings where hosts can collect them on the mucociliary surface of their extensile foot and bring them in contact with their gills (Dufour et al., 2014). This mode of symbiont uptake likely explains why different thyasirid species associate with symbionts belonging to different phylogenetic groups (Rodrigues and Duperron, 2011; Batstone and Dufour, 2016), and why conspecifics of co-occurring *Thyasira* cf. *gouldi* may associate with one of three 16S rRNA phylotypes (Batstone and Dufour, 2016). However, as stated previously, thyasirid species may show a high degree of symbiont specificity despite the extracellular location of their symbionts.

Here, we examine the symbiont populations of *Thyasira* cf. *gouldi* from Bonne Bay (Canada) in greater detail by sequencing fragments of both the 16S rRNA gene and the ribulose-1,5-bisphosphate carboxylase (RuBisCo) gene. The latter gene was chosen because it evolves more rapidly than the 16S rRNA gene and has been shown to be phylogenetically informative in

other studies of chemoautotrophic symbionts (Blazejak et al., 2006; Vrijenhoek et al., 2007). We examine: 1) the relationship between16S rRNA phylotype, RuBisCo phylotype, and host OTU; 2) site specificity of phylotypes at three Bonne Bay sampling locations; 3) temporal patterns in phylotype presence; and 4) relationships between gene phylotype and host size. A more detailed investigation of symbiont gene sequences could reveal site specificity in host-symbiont pairings, temporal changes in symbiont identity within this fjord, or any apparent changes in symbiont identity during the thyasirid's growth. These data should improve our understanding of host-symbiont relationships in these bivalves.

2.2 Materials & Methods

2.2.1 Sample Collection

Thyasirids were collected from Bonne Bay, Newfoundland, Canada on nine occasions between October 2009 and May 2012 (Table A1); permits for field sampling (NL-572-11 and NL-992-12) were obtained from Fisheries and Oceans Canada. Sediment was collected using a Peterson grab (radius = 10.5 cm, length = 30 cm, volume = 0.01 m^3) from three sites within the fjord (Fig. 2.1): Neddy's Harbour, Deer Arm and South East Arm. Thyasirids were retrieved from sediments using a sieve with 1 mm mesh, and symbiotic individuals (*Thyasira* cf. *gouldi* OTU 1 and 2, distinguished by their shell shape; Batstone et al., 2014), were retained.

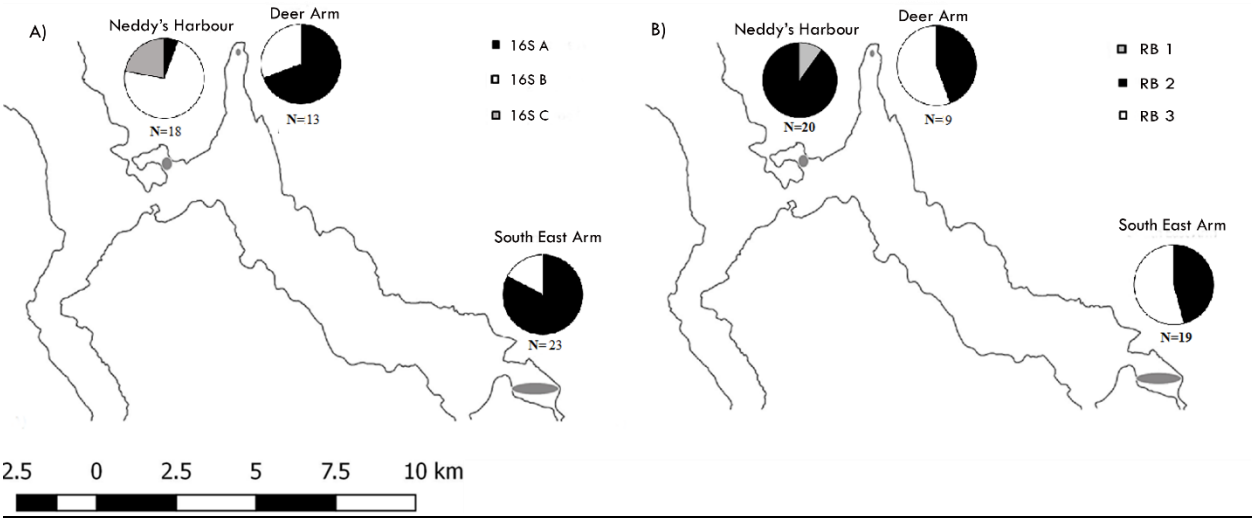

Figure 2.1: Map of sampling sites, 16S rRNA, and RuBisCo phylotypes distributions of *Thyasira* cf. *gouldi* symbionts within Bonne Bay, Newfoundland, Canada.

Sampling sites are located within the shaded ovals. N indicates the number of sequences acquired from each site. A) Distribution of 16S rRNA phylotypes; B) distribution of RuBisCo phylotypes. Adapted from McCuaig et. al, 2017.

2.2.2 DNA Extraction and Gene Sequencing

The gills of symbiotic *Thyasira* cf. *gouldi* specimens were dissected and immediately frozen or stored in 95% ethanol. Following the protocol for animal tissues, total DNA was extracted from gills using QIAGEN DNeasy® Blood and Tissue kit spin columns and stored at -20°C in the elusion buffer provided. The PCR amplification of 16S rRNA and RuBisCo gene sequences was conducted using 12.5 µl of Green Dream Master Mix, 1.5 µl of template DNA, 1 µl of forward primer, 1 µl of reverse primer, and 9 µl of water. A 1323 bp fragment of the 16S rRNA gene was amplified using primers 27F (5'AGAGTTTGGATCMTGGCTCAG 3') and 1492R (5' CGGTTACCTTGTTACGACTT 3') (Lane, 1991). Thermocycler settings were: 94°C for 3 min, 35 cycles of (94°C for 1 min, 50°C for 30 sec, 72°C for 1.5 min) and a final extension of 72°C for 10 min. A 296 bp fragment of the RuBisCo form II gene (previously identified within the symbionts of *Thyasira* cf. *gouldi*; Dufour et al., 2014) was amplified using primers 663F (5' ATCATCAARCTSGGCCTGCGTCCC 3') and

1033R (5' MGAGGTGACSGCRCCGTGRCCRGCMCRTG 3') (Widmer et al., 1998); initial denaturation was at 95°C for 2 min, followed by 30 cycles of 95°C for 1 min, 62°C for 1 min and 72°C for 30 sec, and a final elongation at 72°C for 5 min. PCR products were cleaned using Agencourt AMPure XP Beads (Beckman Coulter, Brea, CA, USA) following the manufacturer's protocol, and sent to The Center for Advanced Genomics, Toronto, Canada for Sanger sequencing.

2.2.3 Phylogenetic Analysis

Sequences were checked for quality, manually trimmed from both ends, and corresponding forward and reverse sequences from a single clam individual were combined into contiguous sequences (contigs) using SEQUENCHER® 5.1 (Gene Codes Corp.). Contigs were then aligned in MEGA 7 (Kumar et al., 2016) using the ClustalW algorithm (Thompson et al., 1994). We paid particular attention to any sites with double peaks in the chromatographs (i.e. where IUPAC degenerate base symbols were assigned by the sequencing software) on corresponding forward and reverse sequences, as these could indicate the presence of more than one gene phylotype within a particular host bivalve; the alignment and corresponding chromatograms were examined closely for such eventualities.

Maximum likelihood phylogenies were constructed for both 16S rRNA and RuBisCo genes with the phylogeny for the former gene including additional sequences from Batstone et al. (2016). Sequences with degenerate bases were not included in the RuBisCo phylogeny as the high number of heterogeneous sequences reduced bootstrap numbers to unacceptable levels. Appropriate nucleotide substitution models were identified using MEGA7 (Figs. 2.2, 2.3). Three distinct clusters were identified in each phylogeny, and representative sequences of the highest quality (and without ambiguities) were chosen for comparison to sequences in the GenBank database (Table 2.1). For each gene, an additional phylogeny was constructed using

representative sequences from the phylotypes identified in this study, along with similar sequences in GenBank and from selected, well-studied chemosymbionts for which both 16S rRNA and RuBisCo form II sequences were available.

The degree of divergence within phylotype groups, and between groups was calculated in MEGA 7, using the Jukes-Cantor model for 16S rRNA and the Tamura-Nei model with 5 discrete gamma distributions for RuBisCo. Variance from the model was estimated with 1000 bootstrap replicates.

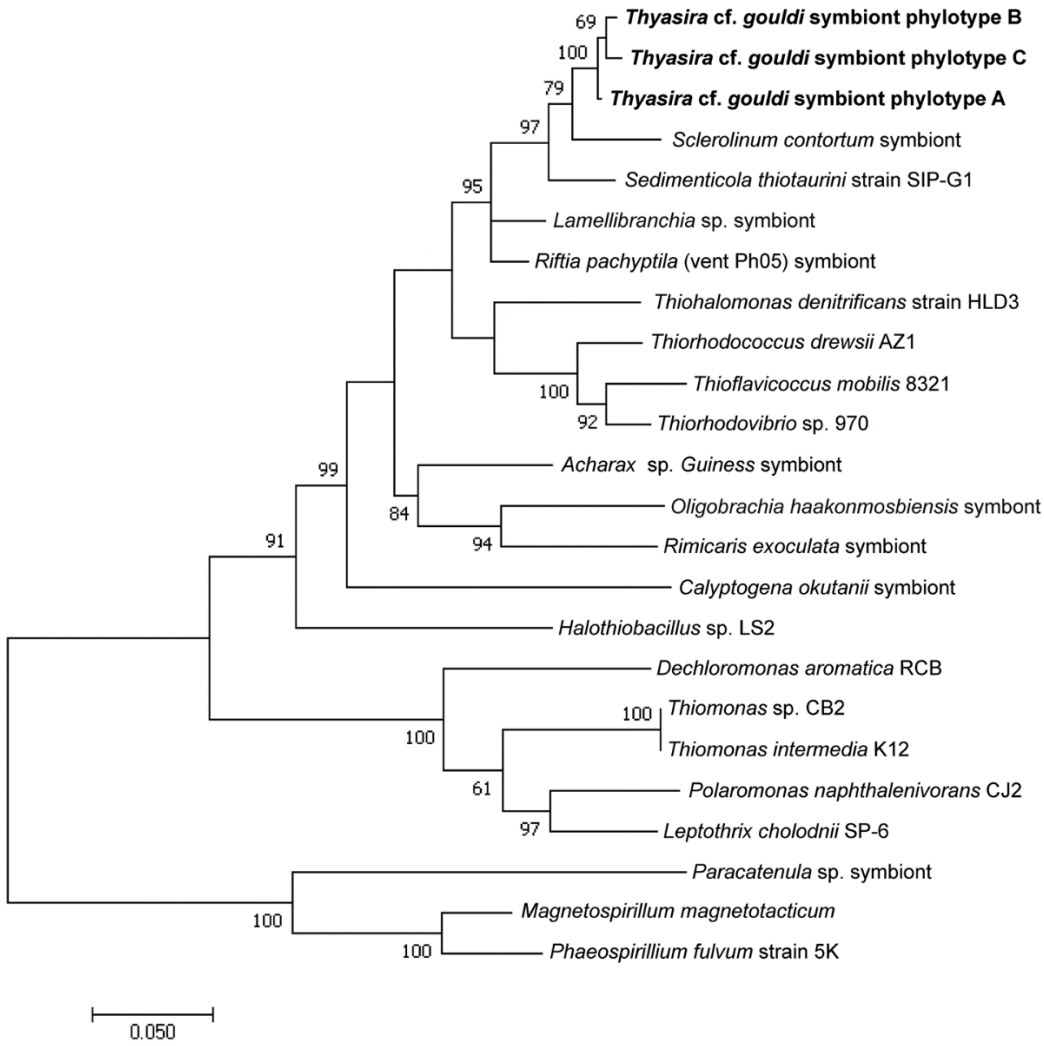

Figure 2.2: 16S rRNA maximum likelihood phylogenies placing *Thyasira* cf. *gouldi* symbiont phylotypes near similar sequences identified in GenBank.

This maximum likelihood phylogeny was created using the Kimura 2-parameter model (Kimura, 1980). The log likelihood of this phylogeny was -5167.5557. A discrete Gamma distribution of 5 was used to model evolutionary rate differences among sites, and the model allowed for some sites to be evolutionarily invariable ([I+] 35.941%). A total of 1323 positions were used in the final dataset. 1000 bootstrap replicates were conducted, with the percentage of phylogenies supporting each branch indicated at branching points. This phylogeny was constructed in MEGA7 (Kumar et al., 2016). Reprinted from McCuaig et. al, 2017.

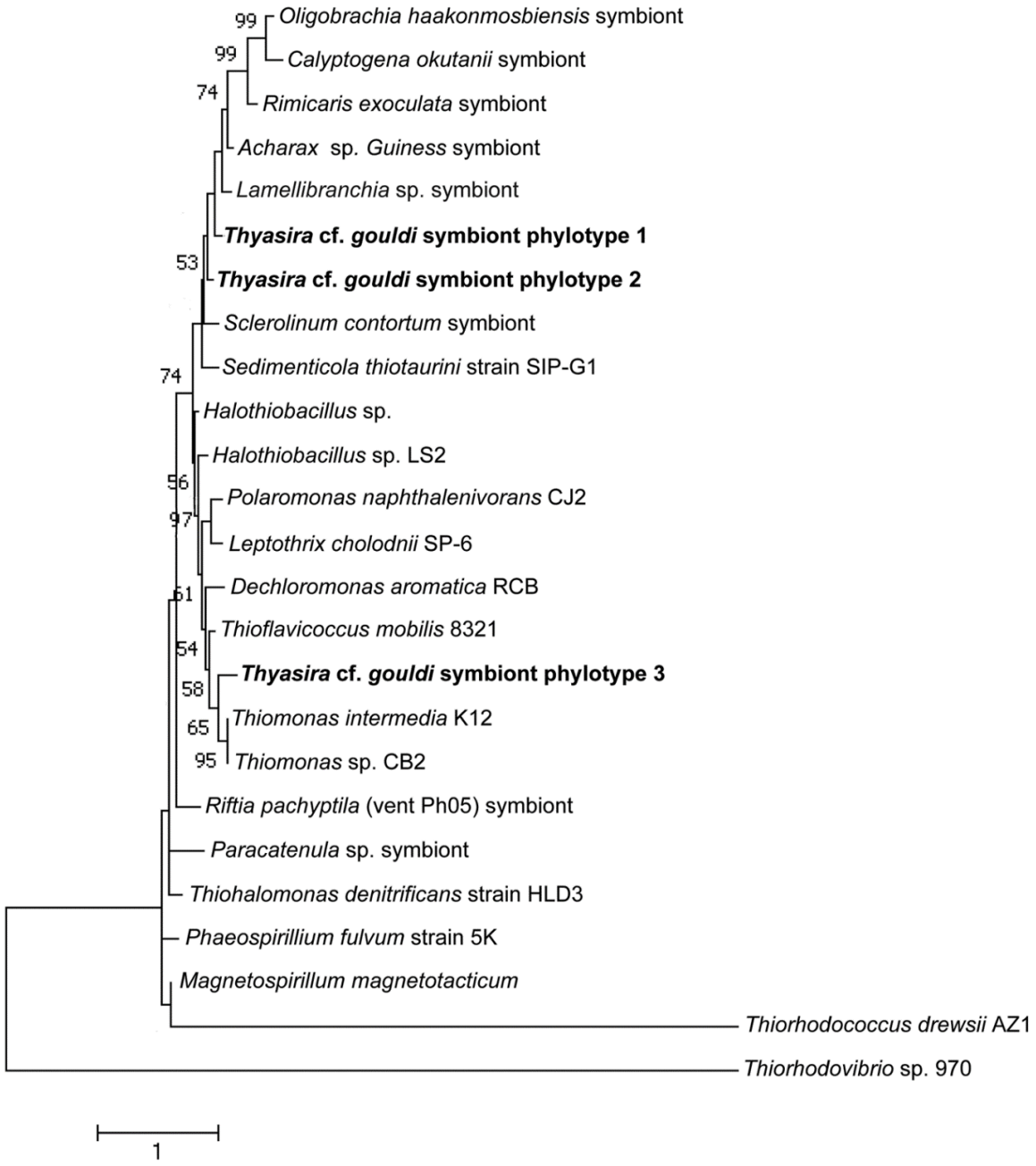

Figure 2.3: RuBisCo maximum likelihood phylogenies placing Thyasira cf. gouldi symbiont phylotypes near similar sequences identified in GenBank

This maximum likelihood phylogeny was created using the Kimura 2-parameter model (Kimura, 1980). The log likelihood of this phylogeny was -5167.5557. A discrete Gamma distribution of 5 was used to model evolutionary rate differences among sites. A total of 296 positions were used in the final dataset. 1000 bootstrap replicates were conducted, with the percentage of phylogneies supporting each branch written at the forks. MEGA7 was used for alignment of sequences and phylogeny construction (Kumar et. al., 2016). Reprinted from McCuaig et. al, 2017.

Table 2.1: GenBank accession numbers of sequences included in the phylogenies.

Organism Name	16S rRNA Accession numbers	RuBisCO Accession Number
Thyasira cf. *gouldi* symbiont phylotype A	MF040754	
Thyasira cf. *gouldi* symbiont phylotype B	MF040755	
Thyasira cf. *gouldi* symbiont phylotype C	MF040756	
Thyasira cf. *gouldi* symbiont phylotype 1		MF040757
Thyasira cf. *gouldi* symbiont phylotype 2		MF040758
Thyasira cf. *gouldi* symbiont phylotype 3		MF040759
Candidatus Vesicomyosocius okutanii HA	AP009247.1	AP009247.1
Bacterium symbiont of *Acharax* sp. Guiness	HE863797.1	HE863799.1
Lamellibrachia sp. endosymbiont	FM165437.1	FM165442.1
Sedimenticola thiotaurini strain SIP-G1	CP011412.1	CP011412.1
Endosymbiont of *Riftia pachyptila*	AY129116.2	AF047688.1
Halothiobacillus sp. LS2	CP016027.1	CP016027.1
Thioflavicoccus mobilis 8321	CP003051.1	CP003051.1
Thiomonas sp. CB2	LK931616.1	LK931649.1
Thiomonas intermedia K12	CP002021.1	CP002021.1

2.3 Results

2.3.1 Description of Sequences and Evolutionary Patterns

From the bivalves examined in this study, we obtained 16S rRNA and RuBisCo form II sequences from 54 and 48 host individuals, respectively. For each gene a phylogeny was constructed using all sequences from this study; for the RuBisCo phylogeny, only unambiguous sequences with no degenerate bases were used. The 16S rRNA sequences formed 3 phylotypes, in agreement with Batstone et al. 2016; these are hereafter referred to as 16S A, 16S B, and 16S C. RuBisCo sequences could also be grouped into three phylotypes: RB1, RB 2 and RB 3.

For each gene, evolutionary distances between and within phylotypes were calculated using the Jukes-Cantor model for 16S rRNA and the Tamura-Nei model with 5 discrete gamma distributions for RuBisCo (Table 2.2). No measurable evolutionary distances were observed within phylotypes of either gene, likely because most discrepancies within RuBisCo phylotypes were within wobble positions. Evolutionary distances between phylotypes were greater for the RuBisCo gene than for 16S rRNA. 16S A and 16S B are more similar to each other (distance =

0.002) than to 16S C (distance = 0.008 and 0.007 respectively). A similar pattern is seen among RuBisCo sequences, with RB 3 being the most distant, as we calculated distance values of 0.229 and 0.267 in comparison to RB 1 and RB 2, respectively. A clear divergence in the evolutionary history of 16S rRNA and RuBisCo genes within the *Thyasira* cf. *gouldi* symbionts was evident upon examination of phylogenies that included sequences from the same free-living bacteria and chemosymbionts (Figs. 2.2, 2.3).

Table 2.2: Evolutionary Distance matrices for 16S and RuBisCo phylotypes.

	16S A	**16S B**	**16S C**		**RB1**	**RB2**	**RB3**
16S A (29)		*0.001*	*0.002*	**RB1** (2)		*0.024*	*0.023*
16S B (21)	0.002		*0.002*	**RB2** (33)	0.131		*0.041*
16S C (4)	0.008	0.007		**RB3** (13)	0.229	0.267	

16S distances calculated using the Jukes-Cantor model with 1000 bootstrap replicates within MEGA 7. RuBisCo distances calculated within MEGA 7 using the Tamura-Nei model and 5 gamma distributions with 1000 bootstrap replicates. The number of sequences (specimens) for each phylotype are in parentheses. Standard Error values in italics.

2.3.2 Evidence for Multiple Symbiont Phylotypes within a Host

Sequences that included degenerate bases were commonly observed (Table A2). Of the 54 16S rRNA sequences, 11 had a single instance of a high-quality call of multiple bases, while 11 others showed 3 to 12 degenerate bases in more variable regions of the gene. Among the 48 RuBisCo sequences, 21 showed high quality calls of multiple bases within variable gene regions. The position of degenerate bases in sequence alignments strongly suggests a mixture of bacterial phylotypes within those host individuals (Table A2).

2.3.3 Phylotypes and Associated Host or Symbiont Characteristics

Our expanded dataset included symbionts from 54 host individuals: 29 with 16S A, 21 with 16S B, and 4 with 16S C symbiont phylotypes (Table A1). We obtained corresponding 16S rRNA and RuBisCo sequences from 34 host specimens and noted patterns in gene pairings within host individuals (Table 2.3). Phylotypes 16S C and RB 1 appeared associated with each other: two of the four host specimens with 16S C had RB 1 (no RuBisCo phylotype data could be obtained from the remaining two host specimens due to insufficient volumes of extracted DNA). Similarly, all 15 hosts containing the 16S B phylotype possessed RB 2. In contrast, seven of the 29 *Thyasira* cf. *gouldi* individuals with symbiont phylotype 16S A had RB 2, while 10 had RB 3.

Table 2.3: Co-occurrence of 16S rRNA and RuBisCo phylotypes within individual *Thyasira* cf. *gouldi* specimens.

	RuBisCo Phylotype		
16S Phylotype	**1**	**2**	**3**
A	0	7	10
B	0	15	0
C	2	0	0

Values are numbers of specimens in which each combination of 16S and RuBisCo phylotype was identified.

Some patterns in the spatial distribution of symbiont 16S rRNA and RuBisCo phylotypes within Bonne Bay were observed. Phylotypes 16S A and B were identified from all the sampling sites. 16S A was more common than B at Deer Arm and South East Arm and was rare at Neddy's Harbour (Fig. 2.1). Phylotypes 16S C (N = 4) and RB 1 (N = 2) were found in the same bivalve specimens, all collected in Neddy's Harbour, suggesting site restriction. RB 3 was found at Deer Arm (N = 5) and South East Arm (N = 8), but not at Neddy's Harbour.

Specimens were collected over a span of four years, and hosts with symbiont phylotypes 16S A, 16S B, RB 2 and RB 3 were identified in all months and years (Table A1). Symbiont phylotypes 16S C and RB 3 were only identified in 2010 and 2011. We found no apparent correspondence between host OTU and symbiont phylotype; the more common OTU 1 associated with all symbiont phylotypes, and the two individuals of host OTU 2 contained symbionts with 16S A/RB 2 and 16S C/RB 1, respectively. Finally, host size showed no obvious relationship with symbiont phylotype (Table A1).

2.4 Discussion

Thyasira cf. *gouldi* hosts a single species of gammaproteobacteria comprising three 16S rRNA subtypes, previously described as phylotypes A, B and C (Batstone and Dufour, 2016). Our phylogenetic investigation revealed a similar level of relatedness and phylogeny topology as reported previously (Batstone and Dufour, 2016). The 16S rRNA symbiont diversity observed in *Thyasira* cf. *gouldi* is unlike that in lucinid clams (Brissac et al., 2016), oligochaetes (Blazejak et al., 2006) and vestimentiferan tubeworms (Vrijenhoek et al., 2007), where a single, invariant symbiont 16S rRNA sequence is present across multiple host species. The evolutionary distances observed among the *Thyasira* cf. *gouldi* 16S rRNA sequences were small (0.002—0.008) with no measurable distance within phylotypes. In contrast, the multiple, co-occurring extracellular symbiont phylotypes observed in bathymodiolins are typically more phylogenetically divergent (Duperron et al., 2008b). This suggests that *Thyasira* cf. *gouldi* may associate with a more restricted group of symbionts than other chemosymbiotic hosts, at least with respect to 16S rRNA gene diversity.

A comparably higher degree of variability and evolutionary distance was observed in the RuBisCo sequences, as was expected. Inter-phylotype distances were greater for the symbiont RuBisCo genes than for 16S rRNA sequences. RB 1 and RB 2 were evolutionarily the closest

(distance of 0.131), while RB 3 was the farthest, with distance metrics of 0.229 from RB 1 and 0.267 from RB 2 (Table 2.2). Similar RuBisCo polymorphism was observed in *Escarpia spicata* and *Lamellibrachia barhami*, but not *Riftia pachyptila* (Vrijenhoek et al., 2007). This higher variability in the RuBisCo gene makes it a useful tool for examining symbiont diversity.

2.4.1 RuBisCo Diversity and Spatial-Temporal Patterns

The greater number of samples analyzed herein has led to a revision of the spatial distribution of 16S rRNA phylotypes since Batstone and Dufour (2016), notably through a greater 16S rRNA phylotype diversity at Deer Arm and South East Arm than previously reported. It is now apparent that 16S A and 16S B are widespread in Bonne Bay, while 16S C has only been identified in Neddy's Harbour. Some RuBisCo phylotypes occurred at multiple sampling sites, but RB1, found only in hosts having 16S C, appeared restricted to the shallower Neddy's Harbour site. The restriction of 16S C/RB 1 to a single site supports the symbiont environmental acquisition mode proposed for thyasirids (Duperron et al., 2012; Dufour et al., 2014).

The ecotype hypothesis proposes that bacterial strains assemble in relation to environmental factors such as sedimentary organic matter, grain size, and sulfur content. Therefore, in environmentally acquired symbionts, the distribution of symbiont phylotypes will reflect these habitat characteristics (Brissac et al., 2016). This hypothesis may explain the apparent absence of RB 3, and the possible restriction of the 16S C/ RB 1 phylotype in Neddy's Harbour, which is the shallowest sampling site, with the lowest organic matter content and coarsest sediments (Batstone and Dufour, 2016). The Deer Arm and South East Arm sites had similar water depths and sediment characteristics (Batstone and Dufour, 2016) that may be conducive to all phylotypes but 16S C/ RB 1. The more widespread RB 2 phylotype may be able to function within a broader range of environmental conditions than RB 1 and RB 3.

Environmental patchiness on small, mm- to cm-scales can explain why thyasirids housing different symbiont phylotypes were found within the same Peterson grab sample (Table A1). At the Bonne Bay sites, a high degree of spatial patchiness in thyasirid abundance and sulfide concentrations was previously noted (Batstone and Dufour, 2016) and may support multiple bacterial strains.

No temporal patterns were recognized within this sample set. We noted no relationship between symbiont phylotype and month or year of collection; the relative rarity of the 16S C/ RB 1 phylotype may explain why it was not identified in all years. We therefore find no evidence for a change in symbiont populations over the span of 2009-2012, although further sampling would be useful to determine this with more confidence.

2.4.2 Symbiont Evolution and the Relationship Between 16S and RuBisCo Phylotypes

The *Thyasira* cf. *gouldi* symbiont16S rRNA and RuBisCo phylogenies are not entirely congruent, although some conservation of gene pairs was identified. While the three 16S rRNA phylotypes are closely related and form a single cluster, the RuBisCo genes in these same symbionts are more evolutionarily dispersed (Fig. 2.3). As noted previously, 16S C was only found in conjunction with RB 1, and 16S B was found in conjunction with RB 2. Phylotype 16S A was found in conjunction with RB 2 and RB 3 in almost equal numbers (Table 2.3). A similar pattern in associated 16S rRNA and RuBisCo gene sequences has been noted in autotrophic proteobacteria and cyanobacteria and attributed to horizontal transfer of the RuBisCo gene amongst different phylogenetic lineages (Delwiche and Palmer, 1996; Elsaied and Naganuma, 2001). During their free-living existence, symbionts would be exposed to other species of bacteria in the sediment and could undergo horizontal gene transfer (HGT) events while outside the host (Dahlberg et al., 1998; Davison, 1999).

The phylogenies highlight the disparity in phylogenetic histories for the 16S rRNA and RuBisCo genes (Figs. 2.2, 2.3). The 16S rRNA phylotypes form a tight cluster near *Sedimenticola thiotaurini* strain SIP-G1, and RuBisCo phylotypes RB 1 and RB 2 are similar to other symbionts (Fig. 2.3). However, RB 3 clusters with free-living bacteria and was likely transferred from another bacterial species at the Deer Arm and South East Arm sites. Its absence from Neddy's Harbour could be due to environmental conditions that exclude these bacteria from this site.

While the slight variation in 16S rRNA genes amongst *Thyasira* cf. *gouldi* symbionts may not be reflective of physiological differences, RuBisCo gene variants may be biologically significant, providing fitness benefits to symbionts under particular environmental conditions. The horizontal transfer of genes in bacterial symbionts may increase their metabolic efficiency, and therefore thyasirid symbioses may be particularly flexible by acquiring symbionts that are locally adapted to their microenvironment. *Bathymodiolus septemdierum*, another bivalve with environmentally acquired sulfur oxidizing symbionts, shows symbiont genomic variation linked with differences in metabolic capabilities, thought to be the result of HGT (Ikuta et al., 2016).

2.4.3 Host-Symbiont Interaction

As observed previously, the *Thyasira* cf. *gouldi* OTUs present in Bonne Bay do not show co-speciation with their symbionts; rather, both host OTUs 1 and 2 can form symbioses with a restricted, but diverse group of bacteria present in the environment (Batstone and Dufour, 2016).

Our data indicate that some *Thyasira* cf. *gouldi* individuals appear to host more than one symbiont phylotype (i.e., they show multi-inoculations, or have mixed symbiont populations). Therefore, the symbiont population is heterogeneous rather than the clonal population found in lucinid bivalves (Brissac et al., 2011; Brissac et al., 2016). Heterogeneous symbiont populations (at both the strain and species level) have been observed in the mytilid species with both

extracellular and intracellular symbionts (Won et al., 2003; Ikuta et al., 2016). The degree of symbiont selectivity appears to vary among chemosymbiotic bivalves, at least in adults.

In lucinids, aposymbiotic juveniles may pick up multiple symbiont strains, which are then maintained in undifferentiated cells dispersed throughout the lateral zone of gill filaments (Brissac et al., 2011). These gill cells later differentiate into mature bacteriocytes (Gros et al., 1997). Theoretically, from this mixed infection, the bacteriocytes with the best energetic yield are kept and imprinted with that bacterial strain throughout their lifetime, coupling the adult host with a specific strain of symbiont (Gros et al., 2012; Brissac et al., 2016). In contrast, *Thyasira* cf. *gouldi* can be co-infected by multiple symbiont strains, even as adults (shell sizes of host clams with the heterogeneous sequences are not smaller than those of clams having no evident symbiont heterogeneity; Table A1). Thyasirids may have the capability of acquiring new and genetically mixed symbionts over their lifetime, potentially increasing metabolic fitness of the holobiont in changing environments, such as temperate and subarctic fjords. The pedal mining behaviour of thyasirids is a possible mechanism for the uptake of new symbionts over the course of the host's life (Dufour et al., 2014, Zanzerl and Dufour, 2017).

2.5 Conclusions

In the Bonne Bay fjord, *Thyasira* cf. *gouldi* individuals can associate with a variety of symbiont phylotypes, and appear capable of hosting mixed symbiont populations, representing a level of host-symbiont specificity that is slightly lower than previously found (Batstone and Dufour, 2016). These findings provide evidence for the horizontal transfer of genetic material between symbionts and free-living bacteria and support the opportunistic environmental acquisition model proposed for this family (Batstone and Dufour, 2016). The capability of *Thyasira* cf. *gouldi* to associate with different strains of bacteria may lead to improved fitness within the variable environment of the fjord sediment and contribute to its phylogenetic diversity

and success in various ecological niches. Some symbiont phylotypes (16S B and RB 2) were found within all sampling sites, while others (16S C and RB 1) appeared restricted to a single site.

This work highlights the importance of looking past 16S rRNA diversity and investigating the functional diversity of symbiont populations. Further work should examine differences between the *Thyasira* cf. *gouldi* symbiont phylotypes in greater detail through genomic or proteomic investigations. Similar work in other chemosymbiotic hosts with extracellular symbionts would also be warranted for a more comprehensive understanding of specificity and selectivity in chemosymbioses.

2.6 Acknowledgements
B Hooper and staff at the Bonne Bay Marine Station assisted with sample collection.

2.7 References

Batstone RT, Dufour SC. 2016. Closely related thyasirid bivalves associate with multiple symbiont phylotypes. *Marine Ecology* 37:988-997.

Batstone RT, Laurich JR, Salvo F, Dufour SC. 2014. Divergent chemosymbiosis-related characters in *Thyasira* cf. *gouldi* (Bivalvia: Thyasiridae). *PLoS ONE* 9:e92856.

Blazejak A, Kuever J, Erséus C, Amann R, Dubilier N. 2006. Phylogeny of 16S rRNA, ribulose 1, 5-bisphosphate carboxylase/oxygenase, and adenosine 5′-phosphosulfate reductase genes from gamma-and alphaproteobacterial symbionts in gutless marine worms (Oligochaeta) from Bermuda and the Bahamas. *Applied and Environmental Microbiology* 72:5527-5536.

Brissac T, Higuet D, Gros O, Merçot H. 2016. Unexpected structured intraspecific diversity of thioautotrophic bacterial gill endosymbionts within the Lucinidae (Mollusca: Bivalvia). *Marine Biology* 163:176.

Brissac T, Merçot H, Gros O. 2011. Lucinidae/sulfur-oxidizing bacteria: ancestral heritage or opportunistic association? Further insights from the Bohol Sea (the Philippines). *FEMS Microbial Ecology* 75:63-76.

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 analysis. *Applied and Environmantal Microbiology* 73:2101-2109.

Cavanaugh C, McKiness ZP, Newton ILG, Stewart F. 2006. Marine Chemosynthetic Symbioses. In: Dworkin M, Falkow SI, Rosenberg E, Schleifer K-H, Stackebrandt E, eds. *The Prokaryotes*. New York, Springer, 475-507.

Dahlberg C, Bergström M, Hermansson M. 1998. In situ detection of high levels of horizontal plasmid transfer in marine bacterial communities. *Applied and Environmental Microbiology* 64:2670-2675.

Dando P, Southward A, Southward E. 2004. Rates of sediment sulphide oxidation by the bivalve mollusc *Thyasira sarsi*. *Marine Ecology Progress Series* 280:181-187.

Davison J. 1999. Genetic exchange between bacteria in the environment. *Plasmid* 42:73-91.

Delwiche CF, Palmer JD. 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Molecular Biology and Evolution* 13:873-882.

Dubilier N, Bergin C, Lott C. 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Reviews Microbiology* 6:725-740.

Dufour SC. 2005. Gill anatomy and the evolution of symbiosis in the bivalve family Thyasiridae. *Biological Bulletin* 208:200-212.

Dufour SC, Felbeck H. 2003. Sulphide mining by the superextensile foot of symbiotic thyasirid bivalves. *Nature* 426:65-67.

Dufour SC, Laurich JR, Batstone RT, McCuaig B, Elliot A, Poduska KM. 2014. Magnetosome-containing bacteria living as symbionts of bivalves. *The ISME Journal* 8:2453-2462.

Duperron S, Halary S, Lorion J, Sibuet M, Gaill F. 2008b. Unexpected co-occurrence of six bacterial symbionts in the gills of the cold seep mussel *Idas* sp. (Bivalvia: Mytilidae). *Environmental Microbiology* 10:433-445.

Duperron S, Laurent MCZ, Gaill F, Gros O. 2008a. Sulphur-oxidizing extracellular bacteria in the gills of Mytilidae associated with wood falls. *FEMS Microbiology Ecology* 63:338-349.

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

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

Goffredi S, Hurtado L, Hallam S, Vrijenhoek R. 2003. Evolutionary relationships of deep-sea vent and cold seep clams (Mollusca: Vesicomyidae) of the "*pacifica/lepta*" species complex. *Marine Biology* 142:311-320.

Gros O, Elisabeth NH, Gustave SDD, Caro A, Dubilier N. 2012. Plasticity of symbiont acquisition throughout the life cycle of the shallow-water tropical lucinid *Codakia orbiculata* (Mollusca: Bivalvia). *Environmental Microbiology* 14:1584-1595.

Gros O, Frenkiel L, Mouëza M. 1997. Embryonic, larval, and post-larval development in the symbiotic clam *Codakia orbicularis* (Bivalvia: Lucinidae). *Invertebrate Biology* 116:86-101.

Hakonen A, Hulth S, Dufour S. 2010. Analytical performance during ratiometric long-term imaging of pH in bioturbated sediments. *Talanta* 81:1393-1401.

Hasegawa M, Kishino H, Yano T-A. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22:160-174.

Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, Aoki Y, Inoue K, Teruya M, Satou K, Teruya K, Shimoji M, Tamotsu H, Hirano T, Maruyama T, Yoshida T. 2016. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. *The ISME Journal* 10:990-1001.

Jones ML. 1981. *Riftia pachyptila* Jones: observations on the vestimentiferan worm from the Galapagos Rift. *Science* 213:333-336.

Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*. 33:1870-1874.

Lane DJ. 1991. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, eds. *Nucleic Acid Techniques in Bacterial Systematics*. Chichester, Wiley, 115-175.

McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America* 110:3229-3236.

Moran NM, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* 42:165-190.

Nussbaumer AD, Fisher CR, Bright M. 2006. Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* 441:345-348.

Ott JA, Novak R, Schiemer F, Hentschel U, Nebelsick M, Polz M. 1991. Tackling the sulfide gradient – a novel strategy involving marine nematodes and chemolithotrophic ectosymbionts. *Marine Ecology* 12:261-279.

Rodrigues CF, Duperron S. 2011. Distinct symbiont lineages in three thyasirid species (Bivalvia: Thyasiridae) from the eastern Atlantic and Mediterranean Sea. *Die Naturwissenschaften* 98:281-287.

Sachs JL, Skophammer RG, Regus JU. 2011. Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* 108:10800-10807.

Stewart F, Cavanaugh C. 2009. Pyrosequencing analysis of endosymbiont population structure: co-occurrence of divergent symbiont lineages in a single vesicomyid host clam. *Environmental Microbiology* 11:2136-2147.

Stewart F, Young CR, Cavanaugh C. 2009. Evidence for homologous recombination intracellular chemosynthetic clam symbionts. *Molecular Biology and Evolution* 26:1391-1404.

Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution* 9:678-687.

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 Research* 22:4673-4680.

Vrijenhoek RC. 2010. Genetics and evolution of deep-sea chemosynthetic bacteria and their invertebrate hosts. In: Kiel S, ed. *The Vent and Seep Biota*. Springer Science+Business Media B.V. 15-49.

Vrijenhoek RC, Duhaime M, Jones WJ. 2007. Subtype variation among bacterial endosymbionts of tubeworms (Annelida: Siboglinidae) from the Gulf of California. *The Biological Bulletin* 212:180-184.

Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. 2014. Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. *Environmental Microbiology* 16: 3699-3713.

Wernegreen JJ. 2005. For better or worse: genomic consequences of intracellular mutualism and parasitism. *Current Opinion in Genetics & Development* 15:572-583.

Widmer F, Seidler RJ, Gillevet PM, Watrud LS, Di Giovanni GD. 1998. A highly selective PCR protocol for detecting 16S rRNA genes of the genus *Pseudomonas* (sensu stricto) in environmental samples. *Applied and Environmental Microbiology* 64:2545-2553.

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*. *Applied and Environmental Microbiology* 69:6785-6792.

Won YJ, Jones WJ, Vrijenhoek RC. 2008. Absence of cospeciation between deep-sea mytilids and their thiotrophic endosymbionts. *Journal of Shellfish Research* 27:129-138.

Zanzerl H, Dufour SC. 2017. The burrowing behaviour of symbiotic and asymbiotic thyasirid bivalves. *Journal of Conchology* 42:299-308.

Chapter 3: Metagenomic analysis suggests broad metabolic potential in extracellular symbionts of the bivalve *Thyasira* cf. *gouldi*

Bonita McCuaig, Lourdes Peña-Castillo, Suzanne C. Dufour

3.0 Abstract

Next-generation sequencing has opened new avenues for studying metabolic capabilities of bacteria that cannot be cultured. Here, we provide a metagenomic description of a chemoautotrophic gammaproteobacterial symbiont population associated with *Thyasira* cf. *gouldi*, a sediment-dwelling bivalve from the family Thyasiridae. Symbionts of thyasirids differ from those of other bivalves by being located outside rather than inside gill epithelial cells, and recent work suggests that they are capable of living freely in the environment. The *T.* cf. *gouldi* symbiont genome shows no signs of genomic reduction and contains many genes that would only be useful outside the host, including flagellar and chemotaxis genes. The thyasirid symbiont may be capable of sulfur oxidation via both the sulfur oxidation and a reverse dissimilatory sulfate reduction pathways, as observed in other bivalve symbionts. In addition, genes for hydrogen oxidation and dissimilatory nitrate reduction were found, suggesting varied metabolic capabilities under a range of redox conditions. The genes of the tricarboxylic acid cycle are also present, along with membrane bound sugar importer channels, suggesting that the bacteria may be mixotrophic. In this study, we have generated the first thyasirid symbiont genomic resources and lay the groundwork for further research in tracking the changes required for life as a bivalve symbiont.

3.1 Introduction

Many species of marine bivalves living near oxic-anoxic boundaries form nutritional symbioses with chemoautotrophic bacteria, which are maintained in or on the host's gills (Stewart et al., 2005; Cavanaugh et al., 2006; Dubillier et al., 2008; Moya et al., 2008). In such

associations, called chemosynthetic symbioses or chemosymbioses, the bacteria provide the host with nutrients and protection from chemical stress, while the host constitutes a protective and suitable environment for the bacterial symbionts (Cavanaugh et al., 2006, Passos et al., 2007, Roeselers and Newton, 2012). The metabolism of symbionts allows hosts to colonize new and often nutrient-poor niches and contributes to their ecological and evolutionary success, as moving into a niche with less competition for resources can lead to evolutionary radiation (Gundus and Douglas, 2009; Stanley, 2014).

Symbionts can be acquired by new generations of hosts in various ways. Vertical transmission is the transfer of bacteria from one generation to the next through gametes, most commonly the eggs; through this pathway, symbionts tend to become obligate. In horizontal transmission, host larvae are inoculated by symbionts released by nearby adults, whereas in environmental transmission, juveniles are inoculated from a free-living symbiont population (Roeselers and Newton, 2012). While the latter mode of transmission does not guarantee symbiont transfer, it does confer some advantages to both partners. Bacteria can avoid genomic reduction (i.e. the deletion of genes that no longer improve symbiont fitness) by maintaining a free-living population. Once genomic reduction occurs, the bacteria cannot survive without the host, and become reliant upon them (Bennet and Moran, 2015); this does not occur in symbionts that maintain a functional environmental population. The maintenance of variation within bacterial populations can also benefit the host, which can be inoculated by symbionts that are well adapted for the local environment, and not necessarily the strain that their parents hosted. As the relationship between symbiotic partners becomes tighter, it may become obligate for both parties. In the case of the host, obligate nutritional symbioses can result in reduction or loss of

the digestive tract, as nutritional reliance upon symbionts increases (Roeselers and Newton 2012).

The symbiont can supplement nutrients that are lacking in the host's diet, or simply provide an additional source of nutrients. The mode of nutrient transfer from symbiont to host varies by relationship, and in many cases, is not well defined. Some symbionts have been shown to actively transfer nutrients to their host, others have "leaky membranes" that allow nutrients to escape the bacteria, and in other cases the host consumes the bacteria through phagocytosis (Felbeck and Jarchow, 1998; Bright et al., 2000). Some metabolic cycles of the symbionts may remove toxins present in the environment, providing the host protection from these compounds (Waite et al., 2008; Podowski et al., 2010). The sox and dsr based sulfur metabolism may remove toxic sulfur compounds while providing energy for carbon fixation. The nitrite reduction (nir) pathway removes harmful nitrogen compounds by using them as an electron sink, but this process is not always coupled with carbon fixation (Liao et al., 2014). One approach to examining the metabolic potential of chemoautotrophic symbionts is to perform genomic, or metagenomic sequencing (Newton et al, 2007; Robidart et al., 2008; Dmytrenko et al., 2014; Konig et al., 2017). By identifying key genes in sequencing data, we can make inferences about the metabolic capabilities of the symbiont.

The bivalve genus *Thyasira* (Family Thyasiridae) contains both symbiotic and asymbiotic species, a seemingly unique condition among bivalve genera (Taylor et al., 2007; Batstone et al., 2014). In contrast to other clams, thyasirids maintain their symbionts among the microvilli of gill epithelial cells, as described in some mussels; such extracellular symbioses have been considered more primitive than intracellular symbioses (Dufour, 2005; Taylor and Glover, 2010; Rodrigues and Duperron, 2011; Roeselers and Newton, 2012). Chemosymbiotic thyasirids are mixotrophs

that appear to rely on particulate food to a greater extent when symbiont abundance is low (Dufour and Felbeck, 2006), or at times when environmental sulfide concentrations are low (Dando and Spiro, 1993). All thyasirid symbionts identified to date are gammaproteobacteria (Dando and Spiro, 1993; Fujiwara et al., 2001; Rodrigues and Duperron, 2011). The thyasirid symbionts are clustered into divergent groups which include both symbiotic and free-living sulfur-oxidizing bacteria (Rodrigues and Duperron, 2011). Enzymatic and PCR techniques have shown the presence of ribulosebisphosphate carboxylase (RuBisCo) and adenylylsulphate reductase in the symbionts of all chemosymbiotic thyasirids investigated (Dando and Spiro, 1993; Rodrigues and Duperron, 2011).

In Bonne Bay, Newfoundland, Canada, gammaproteobacteria have been found living extracellularly on the gills of thyasirid clams identified as *Thyasira* cf. *gouldi* OTUs 1 and 2 (Batstone et al., 2014). In these bivalves, large numbers of symbionts are found in an extensive bacteriocyte zone, with bacteria maintained in extracellular spaces surrounded by the microvilli of bacteriocytes (Batstone et al., 2014). Phylogenetic analysis using 16S rRNA sequences have identified three distinct symbiont phylotypes (A – C) hosted by the two *T.* cf. *gouldi* OTUs (Batstone and Dufour, 2016; McCuaig et al., 2017). There was no apparent co-speciation between host and symbiont as both clam OTUs could host any one of the three symbiont phylotypes, and there is some evidence of multiple strains infecting a single host, although this has not been confirmed (Batstone and Dufour, 2016; McCuaig et al., 2017). The three bacterial 16S rRNA phylotypes cluster during phylogenic analysis and are closely related to the *Thyasira flexuosa* symbiont and to tubeworm symbionts (notably those associated with *Riftia pachyptila*) as well as free-living sulfur oxidizing bacteria (see references Batstone and Dufour, 2016 and McCuaig et al., 2017 for phylogenetic trees). Stable isotope analysis of *T.* cf. *gouldi* supports

chemoautotrophic activity, notably through a lower $\delta^{15}N$ value than in non-symbiotic, co-occurring bivalves; tissue $\delta^{13}C$ values were less negative than in other chemosymbiotic thyasirids, as expected due to the presence of RuBisCo form II (Zanzerl, 2015). *T.* cf. *gouldi* symbionts have been identified within surrounding sediment samples, supporting an environmental mode of transmission and the existence of a free-living symbiont population (Dufour et al., 2014).

We present here the first genomic analysis of a thyasirid symbiont, that of *T.* cf. *gouldi* symbiont phylotype B (one of the most common; Batstone and Dufour, 2016; McCuaig et al., 2017). This investigation is of particular interest given the extracellular location and facultative nature of thyasirid symbionts and provides a contrast to genomic studies of intracellular (and often obligate) bivalve chemosymbionts. After providing an overview of the metagenomic data collected, we characterize important metabolic cycles, including carbon fixation and sulfur oxidation, and identify genomic characteristics that allow us to infer the mode of symbiont transmission and support the evidence for a free-living state in thyasirid bacterial symbionts. By identifying the genes for metabolic pathways in symbiont genomes, we lay the groundwork for future transcriptomic and proteomic studies.

3.2 Methods

3.2.1 Sample Collection and Sequencing

Sediment was collected in August 2010 using a Petersen grab from Neddy's Harbour, in the fjord of Bonne Bay, Newfoundland, Canada (49˚31'21.44"N, 57˚52'11.07"W), at a depth of roughly 15 m under field sampling permits NL 572 11 and NL 992 12 obtained from Fisheries and Oceans Canada. Sediment was wet sieved using a 1 mm mesh and specimens of *T.* cf. *gouldi* were collected and transported to Memorial University, St. John's, Newfoundland. Host individuals were collected in late summer, when symbiont abundance is high (Laurich et al.,

2015). To reduce environmental contamination of bacteria not associated with the gills, the latter were rinsed with 100% ethanol following dissection and before DNA extraction. Total DNA was extracted from the gills of a single individual (host OTU 1; reference Batstone et al., 2014) using a Qiagen Blood and Tissue Kit and stored at -20˚C in the elution buffer provided. Before sequencing, total DNA was transferred to nuclease free water. An Ion Plus Fragment Library Kit (Ion Torrent ^{tm}) was used and fragments of approximately 200 bp were selected using gel size selection and extraction (Qiagen Gel Extraction Kit), purified (Qiagen DNA Purification Kit) following manufacturer's instructions, and concentrations assessed with an Agilent Bioanalyser. Sequencing was conducted on an Ion Torrent PGM Sequencer following the manufacturer's protocols (V2.2). A 316 chip was used for sequencing. Due to poor load rates, two sequencing runs were conducted, and the data were combined before further processing.

3.2.2 Assembly and Annotation

Reads were quality checked and trimmed using FastQC, FAstQ Groomer, FastQ Quality Trimmer, and Filter FastQ the Galaxy Website (usegalaxy.org) and FastQC software (Blankenberg et al., 2010; Afgan et al., 2016); any reads less than 50 bp long were removed at this stage. A quality score of 20 was used, allowing one base below the cutoff score within the read, and trimming was conducted on both ends. Filtered reads were assembled using SPAdes as described below and binned using MEGAN5 (Huson et al., 2007), which used BLAST to compare each contig to the nr database, and all contigs identified as “bacteria”, “unclassified”, “not assigned”, or “no hits" had their reads identified and retained for assembly.

Assembly of the binned data was conducted using SPAdes (Bankevich et al., 2012). Ion Torrent specific settings with kmers 27, 35, 55 and 77 were used. All contigs with lengths of at least 200 bp composed of 8 reads or more were entered into the pipeline. Annotation was run using the MG-RAST website (http://metagenomics.anl.gov/) (Meyer et al., 2008), the RefSeq,

KOG, and Subsystems databases were used with the e-value cut-off set at 5, % identity 60, min length 15, and min abundance 1. A secondary annotation was conducted using PROKKA with the default settings provided (Seeman, 2014).

3.3 Results and Discussion

3.3.1 Genomic overview

Sequence reads can be found on the SRA database under sample SRS1569030, sequencing runs SRR3928943 and SRR3928944. Assembled contigs were uploaded to GenBank under BioProject PRJNA327811, BioSample SAMN05358035, accession numbers MOXF01000001.1-MOXF01012504.1. The assembly resulted in a highly fractured draft genome with similar genes on multiple contigs, suggesting heterogeneity in the bacterial population (recently suggested in reference McCuaig et al., 2017) as genes assembled well, but intergenic spaces did not. Nevertheless, the symbiont population in the *T.* cf. *gouldi* specimen studied was likely comprised of a single species, as only one complete 16S rRNA sequence was found. A similarly fragmented genome despite the presence of a single 16S rRNA sequence was described in a metagenomic study of *R. pachyptila* symbionts, which are environmentally acquired (Robidart et al., 2008).

The assembly resulted in 12,504 contigs, with an N_{50} of 1870. The average read depth coverage is 33, with 50% of bases having a depth coverage of 14 or higher. Bases within areas identified as genes had an average read depth coverage of 31 with 50% having a depth coverage of 20 or higher. The GC content is 42 ± 7%. In total, 20,843 putative genes were assembled. Possible functions were assigned to 3,339 of them allowing us to infer some of the metabolic capabilities of the *T.* cf. *gouldi* symbiont. A summary of Level 1 Subsystem Functions is presented in Table 3.1 below.

Table 3.1: The number of putative proteins assigned to level 1 subsystem functions by the MG-Rast website

Level 1 Subsystem Functions	Number of genes assigned
Amino Acids and Derivatives	211
Carbohydrates	177
Miscellaneous	151
DNA Metabolism	136
Protein Metabolism	148
Cell Wall and Capsule	113
Cofactors, Vitamins, Prosthetic Groups, Pigments	109
RNA Metabolism	107
Regulation and Cell signaling	106
Respiration	102
Membrane Transport	90
Stress Response	66
Virulence, Disease and Defense	61
Nitrogen Metabolism	54
Phages, Prophages, Transposable Elements, Plasmids	42
Fatty Acids, Lipids, and Isoprenoids	40
Motility and Chemotaxis	38
Nucleosides and Nucleotides	38
Sulfur Metabolism	30
Phosphorus Metabolism	25
Cell Division and Cell Cycle	25
Metabolism of Aromatic Compounds	22
Other	382

3.3.2 Genomic support of environmental transmission

Some general characteristics of the genome support the capability of *T.* cf. *gouldi* symbionts to have a free-living existence: there is no sign of genomic reduction (as discussed below). Mobile elements and phage are present in the metagenome. While these are not conclusive evidence of environmental transmission, they are uncommon in co-evolved vertically transmitted symbionts (Moran et al., 2008). Four different mobile elements (Tn10 transposons, ISPsy4, IS200, and a MULE transposase domain possibly from IS256) were identified using the MG-RAST website and PROKKA (see methods), although the exact number of copies was unclear. Phage genes belonging to the order Caudovirales, similar to T5 phages, were also

identified. Genes for phage tail, capsid, and recombinase were identified, although their exact number was again indiscernible.

There was no apparent loss of genes for DNA repair in the *T.* cf. *gouldi* symbiont genome, in contrast to the reduced genome of the vesicomyid symbionts which lacks *rec*A for genetic recombination and *mut*Y for DNA repair (Kuwahara et al., 2011). The loss of genes for DNA repair has been observed in vertically transmitted symbionts, contributing to GC bias and the presence of many pseudogenes (Moran et al., 2008; Kuwahara et al., 2011).

One contig contained many plasmid related genes, and further analysis uncovered the presence of a circular extrachromosomal plasmid. The origin of this extrachromosomal plasmid is uncertain, as this is a metagenomic sample. However, the high number of reads aligned to the plasmid (4860 reads) and complete coverage suggest that it was fairly common within the sample. It is unlikely that a small population of contaminating bacteria would provide enough sequence for a complete plasmid to be assembled, and we therefore consider that it is likely associated with the numerically abundant *T.* cf. *gouldi* symbionts. A *vir*B operon consisting of 10 genes encoding a type IV secretion system was identified on the putative plasmid (Fig. 3.1). The type IV secretion system can be used in conjunction with pili for conjugation, however, the *vir*B operon can also be critical in both pathogenic and mutualistic relationships as a secretion system that moves molecules from bacteria to the host (Dale and Moran, 2006). In many mutualistic relationships, these molecules act to identify, colonize, and communicate with the host in a non-harmful way, with the most common molecules moved across cell walls by this secretion system being DNA (Christie and Vogel, 2000; Christie, 2004). However, in different bacteria the system can transport a number of different molecules; notably, in some pathogenic species (e.g. *Helicobacter pylori* and *Bordetella pertussis*), the system can transfer small effector proteins

such as toxins or virulence factors involved in avoiding the host's immune system (Christie and Vogel, 2000). The genes *vir*B1-5 are often transcribed together and *vir*B7-11 form another co-transcribed group (Christie, 2004). On the plasmid discovered in the *T*. cf. *gouldi* metagenome, the *vir*B genes show a similar arrangement, with two hypothetical proteins placed between *vir*B5 and *vir*B6 (Fig. 3.1). Effector sequences were not identified, and no putative function was found for the nine hypothetical proteins on the plasmid. The importance of this plasmid is unknown, and more investigation is needed to confirm its association with *T*. cf. *gouldi* symbionts and its potential involvement in host-symbiont translocation.

Figure 3.1: Arrangement of the type IV secretion system genes on the extrachromosomal circular plasmid.

The *vir*B T4SS genes are arranged in 2 operons, *vir*B1-*vir*B5 and *vir*B6-*vir*B11, in the same orientation and separated by 2 hypothetical proteins. The *tra*G gene, relaxase/mobilization gene, as well as the DNA-invertase suggest that the T4SS is active as a conjugation system.

The genomic data showed genes associated with flagellar assembly and function (*fla*G, *flg*A, B, C, E, F, G, H, I, J, K, L, *flh*A, B, F, *fli*D, E, G, H, K, L, M, N, P, Q, S, T, U, W, *mcp*B, *mcp*S, *mot*B, *mot*D, *pct*C, *pom*A, *swr*C, *tar*, *ycg*R, and an undefined flagellar motor protein). Also identified were the Che genes (*che*A, B, R, V, W, Y, and Z), which can detect chemical conditions in the environment, and interact with the flagellar motor to help the bacteria move to

suitable areas within the environment (Ferrandez et al., 2002). These genes are essential to locate and move to the microaerobic, reduced sulfur rich areas of the sediment this bacterium needs for sulfur oxidation. An aerotaxis gene (similar to *aer*) was also identified, likely allowing the bacteria to locate the microaerobic areas where sulfur oxidation is most efficiently carried out. When associated with a host, reduced sulfur is made accessible to symbionts by the sulfur mining behavior of the clam; however, bacteria in the free-living population must retain key genes for substrate location and motility (Dufour et al., 2014). Like the environmentally transferred *R. pachyptila* symbiont, the *T.* cf. *gouldi* symbiont has a full complement of flagellar genes, as well as an array of chemotaxis genes (Robidart et al., 2008). Surprisingly, no magnetotaxis genes were identified by our metagenomic analysis, although magnetosome particles were identified in the symbionts of *T.* cf. *gouldi* (Dufour et al., 2014). A directed BLAST search of our contigs using the known magnetosome genes *mam*, *man*, *mms*, and *mad*, also gave no results. Our inability to identify magnetosome genes could be due to the fact that there is little available information on magnetosome genes in gammaproteobacteria, in contrast to other lineages of magnetotactic bacteria. Alternatively, the fragmented nature of the genome may mean they were not found because they fell within gaps of the assembly.

A schematic representation of inferred metabolic capabilities of the *T.* cf. *gouldi* symbiont is presented in Fig. 3.2. The symbiont may not be restricted to thiotrophy and may be able to use alternative metabolic pathways when reduced sulfur is not available. In culturing experiments, the sulfur oxidizing bacterium *Sedimenticola thiotaurini* SIP-G1 is unable to fix carbon in aerobic conditions, where it must instead rely on heterotrophy (Flood et al., 2015). A previous phylogenetic study (McCuaig et al., 2017) placed the *T.* cf. *gouldi* symbiont in a position near *S. thiotaurini* SIP-G1. Based on this phylogenetic placement and the genes

identified by this study, the *T.* cf. *gouldi* symbiont may have similar metabolic capabilities; however, without culturing the bacteria in the lab we cannot validate this theory.

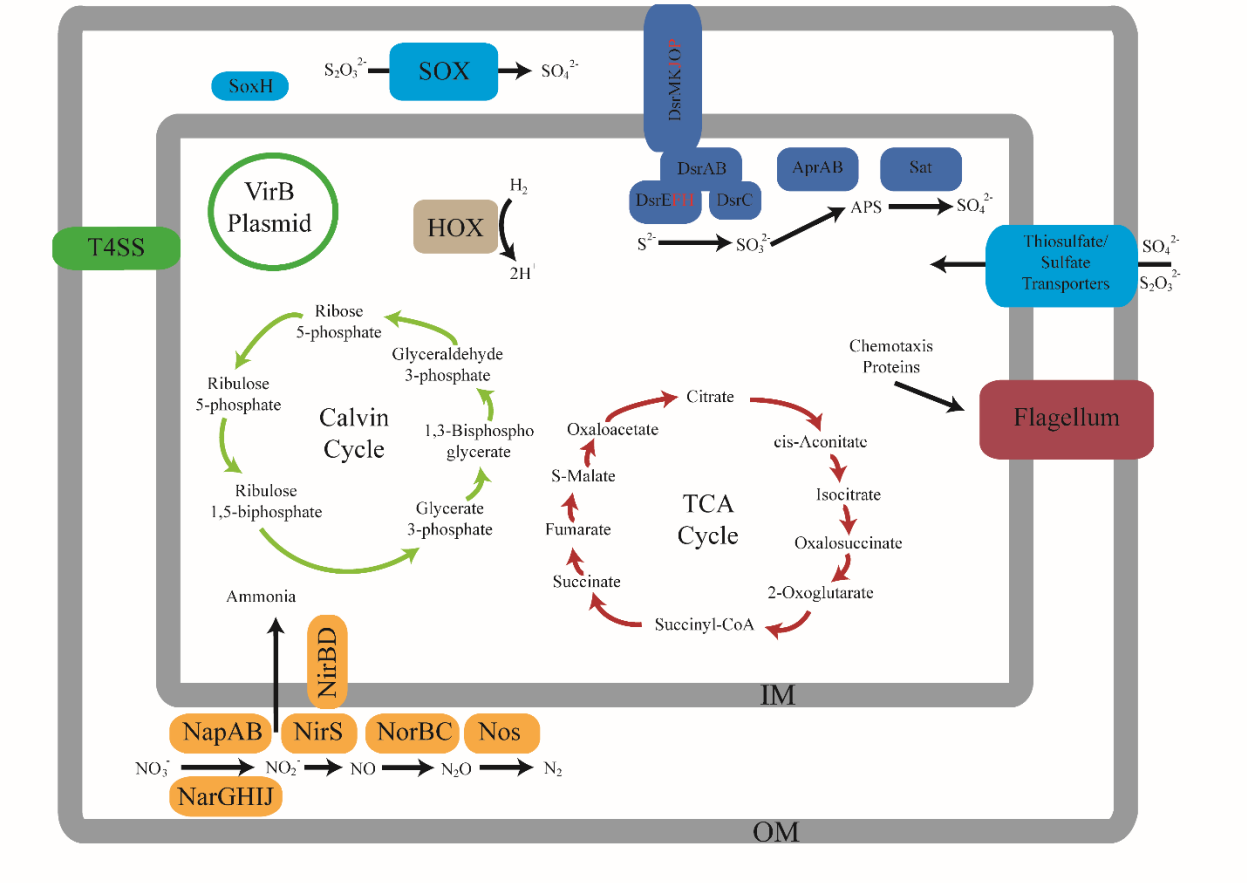

Figure 3.2: Schematic representation of inferred metabolic capabilities of the *Thyasira* cf. *gouldi* symbiont.
Genes for complete CBB and TCA cycles were identified. Genes that were not identified within either annotation pipeline are in red. IM: inner membrane; OM: outer membrane.

3.3.3 Amino acid and cofactor synthesis

Symbionts commonly retain genes important for amino acid, vitamin, and cofactor production because the host selects for bacteria that provide the nutrients it requires (Newton et al., 2007). Many putative gene functions of the *T.* cf. *gouldi* symbiont are involved in amino acid transport and metabolism (229 assignments; Table 3.1), while cofactor transportation and metabolism are also frequently identified (112 times; Table 3.1). These functions are also present

in free-living bacteria, so while important for the symbiosis, they are also presumably essential to the bacteria outside the host.

3.3.4 Thioautotrophy

In the *T.* cf. *gouldi* symbiont, the metabolic cycles for carbon fixation and sulfur oxidation are of particular interest. Several genes for the sox and dsr pathways are present (see below), and the symbiont may conduct sulfur oxidation through both these pathways. Both these cycles have been found to function simultaneously in bivalve and vestimentiferan chemosymbionts (Harada et al., 2009; Roeselers et al., 2010; Li et al., 2018). Sulfur compounds within the benthic sediment are patchy, and therefore being able to utilize different forms would increase the habitat range for these bacteria and their bivalve hosts.

*Sox*A, X, Y and Z, which are found in a cluster in the genome of vesicomyid symbionts (Harada et al., 2009) and form a multi-enzyme system that can oxidize various forms of reduced sulfur (sulfide, thiosulfate, elemental sulfur and sulfite) to sulfate (Friedrich et al., 2001; Harada et al., 2009), were found in the *T.* cf. *gouldi* metagenome. We found no evidence for soxCD, which is found in some sulfur-oxidizing bacteria but is lacking in others (including in the *Calyptogena* symbiont; Harada et al., 2009). The lack of soxCD can manifest itself by the presence of bacterial sulfur globules, which appear as white inclusions in transmission electron micrographs of *T.* cf. *gouldi* symbionts, due to sulfur removal during processing (e.g. Fig. 2B of reference Batstone et al., 2014). The *T.* cf. *gouldi* symbiont metagenome included *sox*H, a peripheral, thiosulfate inducible sox gene that is located in the periplasm but is not essential for growth on thiosulfate and has an unknown function (Rother et al., 2001). We also identified *cys*A, shown to import both sulfate and thiosulfate from the environment (Sirko et al., 1990). Adenylylsulfate reductase was found, and its activity was previously detected in thyasirid symbionts (Dando and Spiro, 1993; Rodrigues and Duperron, 2011).

Many of the genes in the dsr cycle were found, *dsr*A, B, and C as well as the peripheral *dsr*E suggesting that the pathway is running in an oxidative direction (Bradley ar al., 2011). These genes as well as *dsr*K, M, R, S are present in the symbiont genome. An oxidative dsr pathway is present in many well-studied symbionts, including those associated with multiple *Calyptogena* species, *R. pachyptila*, and *Crysomallon squamiferum* (Robidart et al., 2008; Harada et al., 2009; Nakagawa et al., 2014).

Thirteen putative functions associated with the Calvin-Benson-Bassham Cycle were discovered in the *T.* cf. *gouldi* symbiont. The Calvin-Benson-Bassham Cycle in the *T.* cf *gouldi* symbiont utilizes a form II RuBisCo enzyme (Batstone et al., 2014; McCuaig et al., 2017). Chemosymbionts of bivalves often have a reversible pyrophosphate-dependent phosphofructokinase in place of the sedoheptulose-1,7-bisphosphatase that this enzyme replaces, and the fructose 1,6 bisphosphatase genes, which are employed in a reverse TCA cycle (Roeselers et al., 2010; Dmytrenko et al, 2014). However, we were unable to identify any of these three genes in our analysis, but did find ribose 5-phosphate isomerase, which is used in the typical Calvin-Benson-Bassham pathway but is missing in the symbionts of *Calyptogena magnifica* and *R. pachyptila* (Newton et al., 2007; Robidart et al., 2008). It is not clear if the thyasirid symbiont has a traditional Calvin-Benson-Bassham cycle, or if the modifications common in other sulfur oxidizing symbionts are also present in this symbiont (Dmytrenko et al, 2014; Roeselers et al., 2010).

3.3.5 Hydrogen Oxidation

The symbiont also appears capable of hydrogen oxidation using the NAD^+-reducing hydrogenase *hox*HYUF, and a second set of closely related genes identified as coding the alpha, beta, delta, and gamma subunits of *hox*S. The enzyme produced by these complexes is bi-directional. It has been described previously in the symbiont of some vestimentiferan worms (Li

et al., 2018; Reveillaud et al., 2018) as well as free-living *Sedimenticola selenatireducens* (Narasingarao et al., 2014).

3.3.6 Heterotrophy

Genes associated with a complete tricarboxylic acid (TCA) cycle were identified in the *T.* cf. *gouldi* symbiont metagenome (Fig. 3.2). Interestingly, the TCA cycle in this symbiont does not appear to use the oxoglutarate shunt and contains both the α ketoglutarate dehydrogenase and citrate synthase enzymes which are commonly lost in chemosymbiotic bacteria and cause the loss of heterotrophic abilities (Dmytrenko et al., 2014). All genes for a functional TCA cycle have been found in the chemosymbiont of *Solemya velum*, which may occur outside of its host (Dmytrenko et al., 2014). The genome of the *R. pachyptila* symbiont also encodes a complete TCA cycle and contains evidence for response to carbon compounds in the environment, suggesting that it can survive heterotrophically outside the host (Robidart et al., 2008). Sugar phosphotransferase systems (PTS) were also identified in our dataset. These systems can import sugars from the environment, increasing the evidence for some heterotrophic ability. Sugar PTS were identified for fructose, mannose, galactose, and sucrose, suggesting that these substrates can be acquired from the environment, supplementing carbon fixation. In pure culture, the sediment bacterium *S. thiotaurini* SIP-G1 is unable to grow on sulfur oxidation alone and must be provided with heterotrophic nutrients (Flood et al., 2015). A similar system may exist within the *T.* cf. *gouldi* symbiont, with heterotrophic growth occurring when environmental conditions are unfavorable for carbon fixation. The ability to utilise multiple carbon sources would be very beneficial during a free-living stage, especially in fluctuating environments where sulfur compounds can be scarce.

3.3.7 Anaerobic respiration

The *T.* cf. *gouldi* symbiont appears to be capable of performing denitrification, as the genes for the nar and nos pathways are present in the metagenome. Denitrification is the process that reduces potentially harmful nitrogen compounds (nitrates, nitrites, and nitric oxide) into harmless, inert N_2 through anaerobic respiration. Denitrification may provide multiple advantages to both host and symbiont, in addition to allowing bacterial ATP synthesis. First, by reducing harmful nitrogenous compounds, the bacteria may protect their host from toxic effects. Second, by decreasing the symbiont's oxygen requirements, there is less competition with the host for this limited resource in the thyasirid's endobenthic environment. Third, the pathways could allow the bacteria to respire anaerobically in anoxic sediments, and therefore broaden the organism's free-living niche. Notably, the closely related free-living bacterium *S. thiotaurini* SIP-G1 from salt marsh sediments is capable of anaerobic respiration using nitrate and nitrite but can also grow under hypoxic conditions (Flood et al., 2015). Dissimilatory nitrate respiration genes have also been identified in the symbionts of *Vesicomyosocius okutanii*, *Bathymodiolus thermophilus*, and a number of vestimentiferan tubeworms (Robidart et al., 2008; Kleiner at al., 2012; Ponnudurai et al., 2017; Li et al., 2018).

Recent work has shown some sulfur-oxidizing chemosymbionts can also fix atmospheric nitrogen into bioavailable forms (Konig et al., 2017; Petersen et al., 2016), however, we did not find any evidence of nitrogen fixation genes in the *T.* cf. *gouldi* symbiont. The closely-related free-living bacterium *S. thiotaurini* SIP-G1 does however have a complete nitrogen fixation pathway (Flood et al., 2015).

3.4 Conclusions

The metagenomic data collected corroborates previous data suggesting a facultative relationship between *T.* cf. *gouldi* and its symbionts, with the host clams being inoculated from

the environment (Dufour et al., 2014; McCuaig et al., 2017). The timing of this inoculation during the host's lifespan is still unclear and further research is needed to determine when and for how long the host is competent for symbiont uptake. The symbiont population is a collection of closely related individuals, although the population is not clonal and some variation is present; the apparent heterogeneity in the symbiont population is likely related to the extracellular location of the bacteria, which limits the host's ability to select for unique strains. There is no evidence of genome reduction in these symbionts, and the metagenomic data supports evidence of an environmental (non-symbiotic) habitat. In particular, the presence of a functional flagellum and chemosensory abilities supports the presence of a free-living population, as reported previously (Dufour et al., 2014).

Our data suggests that the metabolic capabilities of the *T.* cf. *gouldi* symbionts are comparable to previously described sulfur oxidizing bacteria. The symbionts may utilize multiple pathways for sulfur oxidation, both sox and dsr, and the CBB Cycle for carbon fixation. The denitrification pathway that is also present would allow for carbon fixation in anaerobic areas; when outside the host, sulfides are predominantly found in micro-oxic areas. Unlike many obligate symbionts, the thyasirid symbiont appears to have a functional TCA cycle and sugar importers allowing it to be heterotrophic. The bacteria may utilize autotrophy or heterotrophy under different conditions, like *S. thiotaurini* SIP-G1 (Flood et al., 2015).

Further research into the thyasirid symbiont genome may be beneficial in tracking the changes required for life as a bivalve symbiont, and experimental studies could reveal whether symbionts are capable of reverting to a non-symbiotic state after they have become associated with their host. The *T.* cf. *gouldi* symbiosis provides a unique opportunity to investigate how symbioses evolve as this appears to be a relatively less derived and interdependent relationship

compared to other bivalve symbioses which are intracellular. More research into the metabolic capabilities of the symbiont and how they interact with the host would provide insights into how this relationship has evolved, and the mechanisms that allow it to be maintained. Comparing the different symbiont phylotypes capable of associating with a single host species would also improve our understanding of this relationship and of the potential benefits of flexible host-symbiont pairings.

3.5 Acknowledgements

This research was funded by Natural Sciences and Engineering Research Council (NSERC) Discovery Grants (RGPIN 386087-2010 and 06548-2015) to SCD, and an NSERC Postgraduate Scholarship – Doctoral to BM.

We would also like to thank the Lang lab, specifically Yunyun Fu, for training and assistance with the Ion Torrent PGM.

3.6 References

Afgan E, Baker D, Van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research* 44(W1):W3-W10.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibeski AD, Pyshin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevner PA. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Bacteriology* 19(5):455-477.

Batstone RT, Dufour SC. 2016. Closely related thyasirid bivalves associate with multiple symbiont phylotypes. *Marine Ecology* 37:988-997.

Batstone RT, Laurich JR, Salvo F, Dufour SC. 2014. Divergent chemosymbiosis-related characters in *Thyasira* cf. *gouldi* (Bivalvia: Thyasiridae). *PLoS ONE* 9:e92856.

Bennett GM, Moran NA. 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proceedings of the National Academy of Sciences* 112(33):10169-10176.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutanko A, Taylor J. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology* 19(10):1-21.

Bradley AS, Leavitt WD, Johnston DT. 2011. Revisiting the dissimilatory sulfate reduction pathway. *Geobiology* 9:446-457.

Bright M, Keckeis H, Fisher CR. 2000. An autoradiographic examination of carbon fixation, transfer and utilization in the *Riftia pachyptila* symbiosis. *Marine Biology* 136:621-632.

Cavanaugh C, McKiness ZP, Newton ILG, Stewart F. 2006. Marine Chemosynthetic Symbioses. In: Dworkin M, Falkow SI, Rosenberg E, Schleifer K-H, Stackebrandt E, eds. *The Prokaryotes*. New York, Springer, 475-507.

Christie PJ, Vogel JP. 2000. Bacterial type IV secretion: conjugative systems adapted to deliver effector molecules to host cells. *TRENDS in Microbiology* 8(8):354-360.

Christie PJ. 2004. Type IV secretion: *Agrobacterium* VirB/D4 and related conjugation systems. *Biochimica et Biophysica* 219-234.

Dale C, Moran NM. 2006. Molecular interactions between bacterial symbionts and their hosts. *Cell* 126:453-465.

Dando PR, Spiro B. 1993. Varying nutritional dependence of the thyasirid bivalves *Thyasira sarsi* and *T. equalis* on chemoautotrophic symbiontic bacteria, demonstrated by isotope ratios of tissue carbon and shell carbonate. *Marine Ecology Progress Series* 92.

Dmytrenko O, Russell S, Loo WT, Fontanez KM, Liao L, Roeselers G, Charma R, Stewart F, Newton ILG, Woyke T, Wu D, Lang JM, Eisen JA, Cavanaugh C. 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(1):1-20.

Dubilier N, Bergin C, Lott C. 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Reviews Microbiology* 6:725-740.

Dufour SC. 2005. Gill anatomy and the evolution of symbiosis in the bivalve family Thyasiridae. *Biological Bulletin* 208:200-212.

Dufour SC, Felbeck H. 2006. Symbiont abundance in thyasirids (Bivalvia) is related to particulate food and sulphide availability. *Marine Ecology Progress Series* 320:185-194.

Dufour SC, Laurich JR, Batstone RT, McCuaig B, Elliot A, Poduska KM. 2014. Magnetosome-containing bacteria living as symbionts of bivalves. *The ISME Journal* 8:2453-2462.
Felbeck H. 1981. Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* 213:336-338.
Felbeck H, Jarchow J. 1998. Carbon Release from Purified Chemoautotrophic Bacterial Symbionts of the Hydrothermal Vent Tubeworm *Riftia pachyptila*. *Physiological Zoology* 71(3):294-302.
Ferrández A, Hawkins AC, Summerfield DT, Harwood CS. 2002. Cluster II che genes from *Pseudomonas aeruginosa* are required for an optimal chemotactic response. *Journal of Bacteriology* 184(16):4374-4383.
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*. *International Journal of Systematic and Evolutionary Microbiology* 65:2522-2530.
Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J. 2001. Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? *Applied and Environmental Microbiology* 67(7):2873-2882.
Fujiwara Y, Kato C, Masui N, Fujikura K, Kojima S. 2001. Dual symbiosis in the cold-seep thyasrid clam *Maorithyas hadalis* from the hadal zone in the Japan trench, western Pacific. *Marine Ecology Progress Series* 214:151-159.
Gündüz EA, Douglas A. 2009. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proceedings of the Royal Society of London B: Biological Sciences* 276(1658):987-991.
Harada M, Yoshida T, Kuwahara H, Shimamura S, Takaki Y, Kato C, Tetsuya M, Miyake H, Maruyama T. 2009. Expression of genes for sulfur oxidation in the intracellular chemoautotrophic symbiont of the deep-sea bivalve *Calyptogena okutanii*. *Extremophiles* 13:895-903.
Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* 17(3):377-386.
Kleiner M, Petersen JM, Dubilier N. 2012. Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. *Current Opinion in Microbiology* 15:621-631.
König S, Gros O, Heiden SE, Hinzke T, Thuermer A, Poehlein A, Meyer S, Vatin M, Mbéguié-A-Mbéguié D, Tocny J. 2017. Nitrogen fixation in a chemoautotrophic lucinid symbiosis. *Nature Microbiology* 2(1):16193.
Kuwahara H, Takaki Y, Shimamura S, Yoshida T, Maeda T, Kunieda T, Maruyama T. 2011. Loss of genes for DNA recombination and repair in the reductive genome evolution of thioautotrophic symbionts of *Calyptogena* clams. *BMC Evolutionary Biology* 11(285).
Li Y, Liles MR, Halanych KM 2018. Endosymbiont genomes yield clues of tubeworm success. *The ISME Journal* 1.
Liao L, Wankel SD, Wu M, Cavanaugh CM, Girguis PR. 2014. Characterizing the plasticity of nitrogen metabolism by the host and symbionts of the hydrothermal vent chemoautotrophic symbioses *Ridgeia piscesae*. *Molecular Ecology* 23(6):1544-1557.
McCuaig B, Liboiron F, Dufour SC. 2017. The bivalve *Thyasira* cf. *gouldi* hosts chemoautotrophic symbiont populations with strain level diversity. *PeerJ* 5:e3597.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A,Wilkening J, Edwards RA. 2008. The metagenomics RAST server - A public resource for the automatic phylogenetic and funcitonal analysis of metagenomes. *BMC Bioinformatics* 9(386).

Moran NM, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* 42:165-190.

Moya A, Pereto J, Gil R, Latorre A. 2008. Learning to live together: genomic insights into prokaryote-animal symbioses. *Nature Reviews Genetics* 9:218-229.

Nakagawa S, Shimamura S, Takaki Y, Suzuki Y, Murakami S-i, Watanabe T, Fujiyoshi S, Mino S, Sawabe T, Maeda T, Makita H. Nemoto S, Nishimura S-I, Watanabe H, Watsuji T-o, Takai K. 2014. Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. *The ISME Journal* 8(1):40-51.

Narasingarao P, Häggblom MM. 2006. *Sedimenticola selenatireducens*, gen. nov., sp. nov., an anaerobic selenate-respiring bacterium isolated from estuarine sediment. *Systematic and Applied Microbiology* 29(5):382-388.

Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC, Fontanez KM, Lau E, Stewart F, Richardson PM, Barry KW, Saunders E, Detter JC, Wu D, Eisen JA Cavanaugh C. 2007. The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* 315:998-1000.

Passos FD, Meserani GdLC, Gros O. 2007. Structural and ultrastructural analysis of the gills in the bacterial-bearing species *Thyasira falklandica* (Bivalvia, Mollusca*). Zoomorphology* 126:153-162.

Petersen JM, Kemper A, Gruber-Vodicka H, Cardini U, van der Geest M, Kleiner M, Bulgheresi S, Mußmann M, Herbold C, Seah BKB. 2016. Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nature Microbiology* 2:16195.

Podowski EL, Ma S, Luther III GW, Wardrop D, Fisher CR. 2010. Biotic and abiotic factors affecting distributions of megafauna in diffuse flow on andesite and basalt along the Eastern Lau Spreading Center, Tonga. *Marine Ecology Progress Series* 418:25-45.

Ponnudurai R, Sayavedra L, Kleiner M, Heiden SE, Thürmer A, Felbeck H, Schlüter R, Sievert SM, Daniel R, Schweder T. 2017. Genome sequence of the sulfur-oxidizing *Bathymodiolus thermophilus* gill endosymbiont. *Standards in Genomic Sciences* 12(1):50.

Reveillaud J, Anderson R, Reves-Sohn S, Cavanaugh C, Huber JA. 2018. Metagenomic investigation of vestimentiferan tubeworm endosymbionts from Mid-Cayman Rise reveals new insights into metabolism and diversity. *Microbiome* 6(1):19.

Robidart JC, Bench S, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, Allen EE, Felneck H. 2008. Metabolic versitilty of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environmental Microbiology* 10(3):727-737.

Rodrigues CF, Duperron S. 2011. Distinct symbiont lineages in three thyasirid species (Bivalvia: Thyasiridae) from the eastern Atlantic and Mediterranean Sea. *Die Naturwissenschaften* 98:281-287.

Roeselers G, Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC, Fontanez KM, Lau E, Stewart F. Richardson PM, Barry KW, Saunders E, Detter JC, Wu D, Eisen JA, Cavanaugh C. 2010. Complete genome sequence of *Candidatus* Ruthia magnifica. *Standards in Genomic Sciences* 3:163-173.

Roeselers G, Newton ILG. 2012. On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. *Applied and Environmental Microbiology* 94:1-10.
Rother D, Henrich H-J, Quentmeier A, Bardischewsky F, Friedrich CG. 2001. Novel genes of the sox gene cluster, mutagenesis of the flavoprotein SoxF, and evidence for a general sulfur-oxidizing system in *Paracoccus pantotrophus* GB17. *Journal of Bacteriology* 83(15):4499-4508.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068-2069.
Sirko A, Hryniewicz M, Hulanicka D, Bock A. 1990. Sulfate and thiosulfate transport in *Escherichia coli* K-12: Nucleotide sequence and expression the *cysTWAM* gene cluster. *Journal of Bacteriology* 172(6):3351-3357.
Stanley SM. 2014. Evolutionary radiation of shallow-water Lucinidae (Bivalvia with endosymbionts) as a result of the rise of seagrasses and mangroves. *Geology* 42(9):803-806.
Stewart FJ, Newton ILG, Cavanaugh C. 2005. Chemosynthetic endosymbioses: adaptations to oxic-anoxic interfaces. *TRENDS in Microbiology* 13(9):439-448.
Taylor JD, Williams ST, Glover EA. 2007. Evolutionary relationships of the bivalve family Thyasiridae (Mollusca: Bivavlia), monphyly and superfamily status. *Journal of the Marine Biological Association of the United Kingdom* 87:565-574.
Taylor JD, Glover EA. 2010. Chemosymbiotic Bivalves. In: Kiel S, editor. *The Vent and Seep Biota. Topics and Geobiography.* 33: Springer Science+Business Media B.V.
Waite TJ, Moore TS, Childress JJ, Hsu-Kim H, Mullaugh KM, Nuzzio DB, Paschal AN, Tsang J, Fisher CR, Luther III GW. 2008. Variation in sulfur speciation with shellfish presence at a Lau Basin diffuse flow vent site. *Journal of Shellfish Research* 27(1):163-168.
Zanzerl H. 2015. The burrowing behaviour and diet of thyasirid bivalves from Bonne Bay, NL. MSc Thesis, *Department of Biology*, Memorial University of Newfoundland.

Chapter 4: Comparison of Metabolic Capabilities of *Thyasira* cf. *gouldi* Symbiont Phylotypes

4.0 Abstract

The small clam *Thyasira* cf. *gouldi* hosts a symbiotic population of bacteria within the gill filaments. These bacteria are sulfur oxidizing gammaproteobacteria, that have been identified as three separate 16S rRNA phylotypes. The previous chapter describes the genomic characteristics of phylotype B, here we compare the gene content of the three metagenomes. All three genomes had a similar number of genes identified (A=2418, B=2574, C=2529) and metabolic capabilities were very similar between genomes. Genes associated with carbon fixation, sulfur oxidation, hydrogen oxidation, flagella and chemotaxis were found in all three metagenomes. Upon closer inspection it was found that conserved genes often showed more variation at the sequence level than the 16S rRNA comparisons suggest. Certain functions, while conserved, were achieved with alternate enzymes in phylotype C. This phylotype also lacked the putative biofilm genes found in phylotypes A and B. The creation of a biofilm on the gill surface may explain how the symbiont excludes other bacteria from colonizing the gills. Overall this data supports our theory of an environmental acquisition of symbionts by *Thyasira* cf. *gouldi* and does not dispute the possibility of heterogenous symbiont populations within a single host.

4.1 Introduction

Symbiotic relationships between chemoautotrophic bacteria and invertebrates from at least seven phyla are well documented within marine environments (Dubilier et al., 2008). Some species of bivalve molluscs are known to host large populations of chemoautotrophic bacteria on or in their gills, and in return the bacteria provide the clam some or most of its required nutrients (Dubilier et al., 2008; Roeselers and Newton, 2012). These populations of bacteria are often

considered to be monocultures; however, exceptions have been found and mixed populations may be more common than once thought (Ikuta et al., 2016; McCuaig et al., 2017). These mixed populations of bacteria are much more likely when the symbionts are acquired from the environment and may provide more environmental flexibility to the symbiotic relationship, allowing the partners to thrive in changing environments (Cavanaugh et al., 2006; Bright and Bulgheresi, 2010; Wentrup et al., 2014).

To date genetic research is sparse on the symbionts of bivalves from the family Thyasiridae, which are, with some exceptions (Fujiwara et al., 2001; Passos et al., 2007), extracellular (Dufour, 2005; Duperron et al., 2013) and environmentally acquired (Dufour et al., 2014). All the symbionts described to date have been from the class Gammaproteobacteria, and all those tested have been capable of sulfur oxidation (Dando and Spiro, 1993; Rodrigues and Duperron, 2011; Duperron et al., 2013). Previous research on *Thyasira* cf. *gouldi* from a fjord in Newfoundland, Canada, has shown they can associate with three strains of gammaproteobacteria (phylotypes A-C; Batstone and Dufour 2006). While sequencing of 16S rRNA PCR products showed only one strain of bacteria being hosted by each individual clam, further investigation showed evidence of heterogeneity in the RuBisCo gene (McCuaig et al. 2017). Diversity in a gene with such an essential function suggested the different strains of symbiont may have different metabolic repertoires. The different RuBisCo phylotypes had divergent sequences, sharing 77%-87% sequence identity (phylotypes 1 and 2 being most similar, and phylotypes 2 and 3 being the most divergent).

An initial investigation of the metabolic capabilities of the *Thyasira* cf. *gouldi* symbiont phylotype B was conducted and previously described (Chapter 3). This phylotype was found to be capable of sulfur oxidation through both the sox and dsr pathways. It appeared to have a

functional TCA cycle and may be able to live heterotrophically as well as through chemoautotrophy. Phylotype B was also found to be capable of hydrogen oxidation and denitrification. The symbionts of *Solemya velum* and vestimentiferan tubeworms were also found to be very metabolically diverse (Robidart et al. 2008; Dmytrenko et al. 2014; Reveillaud et al. 2018). These symbionts are thought to be acquired from the environment by the host, and these diverse metabolic cycles may be an adaptation to highly variable sediment conditions (Reveillaud et al. 2018). The free-living bacteria *Sedimenticola thiotaurini* shows metabolic diversity similar to phylotype B, described in Chapter 3, and is within the top BLAST hits for multiple genes within the metagenome, including 16S rRNA (McCuaig *et al.* 2017). *Sedimenticola thiotaurini* have been cultured in the lab and employ different metabolic strategies under different environmental conditions. For example, the bacteria are heterotrophic in aerobic conditions, but rely on chemoautotrophy and sulfur oxidation under anaerobic conditions (Flood *et al.* 2015). The symbiont *Endoriftia persephone*, the endosymbiont of *Riftia pachyptila*, has also been shown to utilize different genes when living outside or inside the host (Robidart *et al.* 2008).

Here, we compare the metabolic capabilities of the three identified strains of the *Thyasira* cf. *gouldi* symbiont (phylotypes A, B, and C) using metagenomic data. We also compare the phylogenetic distance of key genes within metabolic cycles, to elucidate the relatedness of these different symbiont phylotypes.

4.2 Methods

4.2.1 Sample Collection and Sequencing

Three specimens of *Thyasira* cf. *gouldi* were examined herein, including two that were collected from Neddy's Harbour, Bonne Bay, Newfoundland, Canada. The first specimen was collected from South East Arm (Bonne Bay) in May 2012, as described in Batstone et al. 2014

and McCuaig et al. 2017. Previously identified samples collected from Neddy's Harbor (Bonne Bay) in August 2010 for a previous study (Batstone & Dufour 2016) were then selected to ensure all 3 phylotypes were represented.

Total DNA was extracted from the gills of single individuals (Batstone et al. 2014) using a Qiagen Blood and Tissue Kit and stored at -20˚C in the elution buffer provided. Before sequencing, total DNA was transferred to nuclease free water. An Ion Plus Fragment Library Kit (Ion Torrent ^{tm}) was used and fragments of approximately 200 bp were selected using gel size selection and extraction (Qiagen Gel Extraction Kit), purified (Qiagen DNA Purification Kit) following manufacturer's instructions, and concentrations assessed with an Agilent Bioanalyser. Sequencing was conducted on an Ion Torrent PGM Sequencer following the manufacturer's protocols (V2.2). A 316 chip was used for sequencing phylotypes B and C, a 318 chip was used for phylotype A. Due to poor load rates, two sequencing runs were conducted when sequencing phylotype B and the data were combined before further processing.

4.2.2 Assembly and Annotation

Reads were quality checked and trimmed using the Galaxy Website (usegalaxy.org) and FastQC software (Blankenberg *et al.* 2010). A quality score cutoff of 20 was used, allowing one base below the cutoff score within the read, and trimming was conducted on both ends. Any reads less than 50 bp long were removed.

Two approaches were used for assembly and annotation. First, the reads were assembled into contigs using SPAdes (Bankevich *et al.* 2012) with settings for Ion Torrent and kmers of 27, 35, 55 and 77. These contigs were annotated using MG-Rast, with default settings.

In the second pipeline, with PROKKA annotation, the quality control steps outlined above were also conducted, but contigs were binned using MEGAN5 (Huson *et al.* 2007). Reads from all bacterial, unknown, and unassigned contigs were removed and used for the second

assembly. We used SPAdes (Bankevich *et al.* 2012) to assemble the binned data, using settings for Ion Torrent and kmers of 27, 35, 55 and 77. Annotation of contigs 200 bp or longer was conducted using the PROKKA software with default settings (Seemann 2014). Comparison of gene content across symbiont phylotypes was done manually after the PROKKA assembly using gene lists with duplicates removed.

4.3 Results and Discussion

4.3.1 Metabolic Similarities

Assembly resulted in highly fragmented genomes, similar to the previously described phylotype B genome (Chapter 3), however, many genes were successfully assembled. The PROKKA and MG-RAst annotations are compared in (Table 4.1). We will focus on the PROKKA annotation pipeline results for most of this chapter, as the results of this pipeline were more easily compared between genomes.

Table 4.1: Comparison of the output of the two annotation pipelines used here.

	Predicted protein features	**# of Identified Functional Categories**
MG-RAst Phylotype A	33,669	3,239
MG-RAst Phylotype B	20,843	3,339
MG-RAst Phylotype C	20,614	4,057
	Predicted Coding Sequences	**# of unique gene functions identified**
PROKKA Phylotype A	33,317	2,418
PROKKA Phylotype B	18,677	2,574
PROKKA Phylotype C	14,183	2,529

The three genomes contained roughly the same number of unique, identified genes: A=2418, B=2574, C=2529. Of these, 1560 were found in all three symbiont genomes (Fig. 4.1). A large proportion of these shared genes had functions related to DNA synthesis and repair, protein synthesis, cellular replication, and other essential functions. Alongside the essential genes, many metabolic genes were also conserved, Table B1. Assuming that the host selects

bacterial symbionts based on function, these conserved metabolic functions are likely important to the symbiotic relationship (Petersen *et al.* 2012). Genes involved in sulfur oxidation, hydrogen oxidation, denitrification, and carbon fixation were all conserved throughout the genome annotations. The possible benefits to the host are discussed below. Metabolic capabilities that are less clearly beneficial to the host were also conserved, including a functional TCA cycle, flagella, and chemotaxis genes; however, these genes would provide benefits to the bacteria when it lives outside the host.

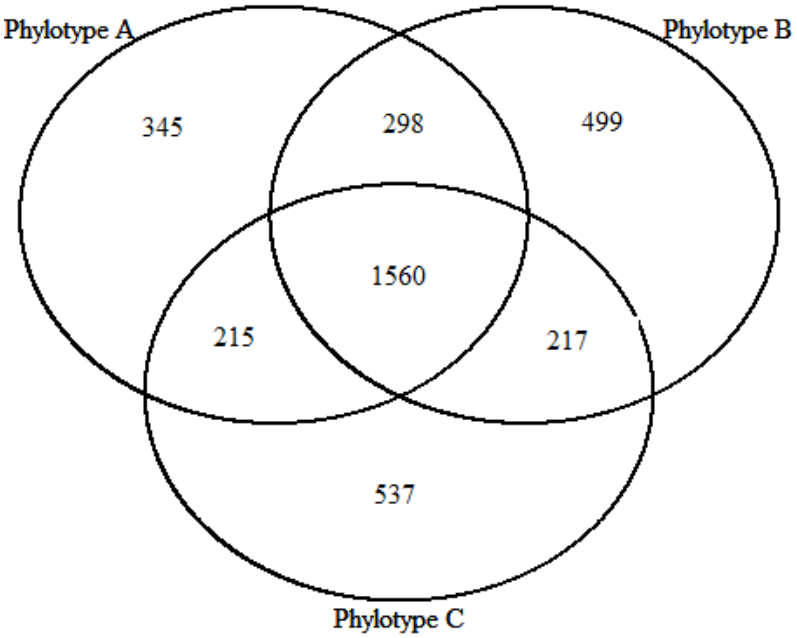

Figure 4.1: Venn diagram of unique genes within each *Thyasira* cf. *gouldi* symbiont phylotype as identified by PROKKA

Symbiont phylotype B was previously described to have both the bacterial sulfur oxidation (sox) pathway and a dissimilatory sulfite reductase (dsr) pathway that ran in the oxidative direction (Chapter 3); these pathways were also identified in phylotypes A and C. This is consistent with research conducted on other thyasirid symbionts, as all those examined to date have been capable of sulfur oxidation, however, in most cases the exact pathway is unknown

(Duperron *et al.* 2013). Genes associated with the Calvin-Bensen-Bassham (CBB) Cycle were present in all three assemblies. Three phylogenetically distinct form II RuBisCo genes were previously described in these symbionts, which were not directly linked to symbiont 16S rRNA phylotype (McCuaig et al., 2017). Each distinct RuBisCo phylotype was represented in one of the genomes discussed here, 16S rRNA phylotype A contained RuBisCo phylotype 3, 16S rRNA phylotype B contained RuBisCo phylotype 2, and 16S rRNA phylotype C contained RuBisCo phylotype 1.

Table 4.2: Distance Matrix for Citrate Synthase sequences in the *Thyasira* cf. *gouldi* symbionts

	Phylotype A Citrate Synthase 1	Phylotype B Citrate Synthase 1	Phylotype C Citrate Synthase 1
Phylotype B Citrate Synthase 1	0.209		
Phylotype C Citrate Synthase 1	0.171	0.175	
Phylotype C Citrate Synthase	0.647	0.625	0.599

The number of base substitutions per site from between sequences are shown. Analyses were conducted using the Tamura-Nei model (Tamura & Nei 1993). The rate variation among sites was modeled with a gamma distribution (shape parameter = 5). All positions containing gaps and missing data were eliminated. There were a total of 1044 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.* 2016).

Full TCA cycles were also found in all three genomes, suggesting the bacteria may also be able to live heterotrophically. The class 1 fumarate hydratase found in phylotype C was annotated as being aerobic, while those in phylotypes A and B were anaerobic. Phylotype C also contained a second, phylogenetically distinct citrate synthase (Table 4.2), and contained a 2-methylisocitrate dehydratase instead of an aconitate hydratase. Sugar importer genes were also identified within all the symbiont genomes, increasing the evidence the bacteria are also capable of living heterotrophically. Symbionts that maintain an environmental population, such as the

Endoriftia persephone and the *Solemya velum* symbiont have also been shown to be mixotrophs (Robidart *et al.* 2008; Dmytrenko *et al.* 2014). The *Thyasira* cf. *gouldi* symbiont also clusters near free-living mixotrophic bacteria such as the genus *Sedimenticola* based on 16S rRNA sequence data (McCuaig et. al., 2017).

Hydrogen oxidation pathways were found in all three genomes. This included two very similar groups of genes, *hoxHYUF* and all four subunits (α, β, γ, and δ) of the *hoxS* complex. This redundancy in all the symbiont strains suggests the function is important to the organisms, and whether this importance is during free-living or symbiotic stages remains unclear. Similar genes were identified in some vestimentiferan worm symbionts (Reveillaud *et al.* 2018). Another similarity with vestimentiferan symbionts is the presence of multi-drug efflux proteins and ABC transporters, which were suggested to potentially function as defence against the host (Reveillaud *et al.* 2018).

All the symbionts also contained *nar*, *nos*, and *nir* genes, which have all been identified in vestimeniferan worm symbionts (Markert *et al.* 2011; Gardebrecht *et al.* 2012; Reveillaud *et al.* 2018). It is possible the bacteria can function in both aerobic and anaerobic environments, which is consistent with the inferred free-living infaunal habitat: the bacteria are believed to live at the sedimentary oxic/anoxic interface, with the burrowing and burrow irrigation behaviours of the clam possibly attracting symbionts to oxic/anoxic interface zones along burrow linings (Dufour et al., 2014). Genes associated with identifying the conditions found at redox boundaries and controlling aerobic and anaerobic functions were also identified (*aer*, *arc*B, *nre*B).

Genes for chemotaxis and motility via flagella were also present in all three genomes. This supports previous work suggesting the symbionts retain a free-living population (Dufour *et al.* 2014). These genes become unnecessary and are often not identified in symbionts that do not

have a free-living stage, such as the vertically transmitted *Calyptogena okutanii* symbiont (Kuwahara *et al.* 2007), but have been identified in the environmentally transferred symbiont *Endoriftia persephone* (Robidart *et al.* 2008).

The MG-RAst analysis also showed some biofilm formation and quorum sensing genes, Table 4.3. If the bacterial symbionts are able to produce a biofilm on the gills of the host, similar to the biofilm formed in the bioluminescence crypts of the bobtail squid (Shibata *et al.* 2012), this may protect them from the host's immune system and help them exclude non-symbiotic bacteria. However, these genes were only identified by the MG-RAst analysis and this was not confirmed by the PROKKA analysis, although other polysaccharide synthesis and transport genes were identified by the latter.

Table 4.3: Biofilm formation and Quorum Sensing Genes Identified by MG-RAst

Gene	Putative Function	MG-RAst Assemblies Where Identified
SypH	Glycosyltransferase	B
SypR	Sugar transferase involved in lipopolysaccharide synthesis	A, B
SypM	Acetyltransferase	B
SypC	Periplasmic protein involved in polysaccharide export	A, B
SypO	Polysaccharide biosynthesis chain length regulator	A, B
SypG	Sigma-54 dependent transcriptional regulator	A, B
SypB	Outer membrane protein	A
RhlL	N-acyl-L-homoserine lactone synthetase	C

The *syp* genes were first described in the symbiont *Vibrio fischeri*, which colonizes the bobtail squid from an environmental population. The *syp* genes were identified in an 18 gene locus (*syp*A-*syp*R), and the regulatory protein rscS was found in another area of the genome (Shibata *et al.* 2012). In the squid-*Vibrio* symbiosis, colonization was not possible following mutations of 16 of the 18 *syp* genes (Shibata *et al.* 2012). The *Thyasira* cf. *gouldi* symbionts may

have analogous genes that were too divergent at the sequence level to be identified or may be able to function with fewer surface proteins. The *syp* genes were only identified in phylotypes A and B, which are the more common symbiont phylotypes (McCuaig *et al.* 2017). Genes that may be associated with host-symbiont recognition were also found to be common among the symbionts. While the *Thyasira* cf. *gouldi* is extracellular, surface proteins may help the host identify symbiotic bacteria and distinguish them from other environmental bacteria (Shibata *et al.* 2012). Two genes associated with host invasion (invasin, and invasion associated locus B (IalB)) were identified in all three genomes. Invasin was identified as an important protein for enteric bacteria to enter mammalian cells (Isberg *et al.* 1987). IalB was identified in all three genomes, but there was no apparent locus A identified in any annotation. This gene is also involved in cell invasion, and has been shown to respond to changes in environmental cues (Coleman & Minnick 2001, 2003). The effect these genes have in this relationship is unclear. The bacteria do not invade the host cells (Batstone *et al.* 2014; Laurich *et al.* 2015), but perhaps they have some function in cell attachment or host recognition.

4.3.2 Phylogenetic Diversity

Variations among the three genomes were observed. Some of the annotated genes were uniquely identified in a particular phylotype (A= 14.3%; B= 19.4%; C=22.2% unique genes). Unlike the previously described phylotype B (Chapter 3), phylotypes A and C did not seem to contain a plasmid, or a type IV secretion system.

Upon closer examination, some of the genes with conserved functions and annotations were phylogenetically distinct (see appendix B). It is also possible that there are multiple strains of symbiont present in a single *Thyasira* cf. *gouldi* host. Attempts to bin out these multiple strains were unsuccessful: binning was attempted using GC content versus length of contigs, GC content versus read depth, and contig length versus read depth, all of which yielded one

continuous bin. Divergent genes with duplicate functions were identified (such as citrate synthase and citrate synthase 1 in phylotype C, see Table 4.2). It is possible that these genes could be present in a single bacterial genome; however, a previous study on the RuBisCo sequences also suggested multiple strains of symbiont within a bacterial population inhabiting a single host (McCuaig, 2017). While some bivalves have been shown to host multiple symbionts, these are phylogenetically diverse species, often utilizing different chemicals for chemoautotrophy (such as sulfide and methane) (Fujiwara *et al.* 2001; Duperron *et al.* 2009), very similar symbionts with diverse metabolic capabilities have been described in *Bathymodiolus septemdierum* (Ikuta *et al.* 2016). These multiple, highly similar strains would also explain some of the difficulty experienced in assembly. Even genes that were present in all three genomes showed phylogenetic variation. Different genes showed variation in the amount of phylogenetic distance among phylotypes. Distance matrices for 16S rRNA and RuBisCo were previously reported (Table 2.2). The 16S rRNA sequences are highly similar, but selected proteins from metabolic cycles show a much higher degree of distance (Tables B2-B4, Appendix B). This suggests that, while functions are highly conserved, the symbionts are not as closely related as the 16S rRNA analysis would suggest.

We also found multiple instances of genes with multiple copies in all three assemblies. One such gene was the denitrification regulatory protein NirQ, Table 4.4. Each genome contained 2 highly divergent copies of *NirQ* (genetic distance >0.67), but the distance between genomes within similar copies was similar to the other genetic distances reported for metabolically important genes (0.123-0.204). This may be a sign of functional redundancy in all symbiont genomes, or the divergent gene copies may function well under different

environmental conditions. The third possibility is that each metagenome was contaminated by similar, but divergent environmental bacteria.

Table 4.4: Distance matrix for Denitrification regulatory protein NirQ in *Thyasira* cf. *gouldi* symbionts

	Phylotype A Copy 1	Phylotype B Copy 1	Phylotype C Copy 1	Phylotype A Copy 2	Phylotype B Copy 2
Phylotype B Copy 1	0.204				
Phylotype C Copy 1	0.201	0.181			
Phylotype A Copy 2	0.671	0.656	0.652		
Phylotype B Copy 2	0.686	0.683	0.669	0.191	
Phylotype C Copy 2	0.673	0.670	0.682	0.186	0.123

*Grey area is distance between copies within the same genome.
Analyses were conducted using the Tamura 3-parameter model with the rate variation modeled with a gamma distribution (shape parameter =5) (Tamura 1992). There were a total of 802 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.* 2016).

4.4 Conclusion

The genomic data obtained in this study shows that the *Thyasira* cf. *gouldi* symbiont has a large range of metabolic capabilities. Sulfur oxidation may be carried out through the sox pathway, or a dsr pathway running in reverse. The bacteria may also possess the capability to live heterotrophically and has a suite of genes for chemotaxis and movement within the environment. They also appear to have the full suite of genes needed for denitrification and hydrogen oxidation. This metabolic flexibility has been seen in environmentally transferred symbionts, and further studies would be needed to identify which pathways are active within the host, and which are only active in the environment. The overall metabolic capacity of the symbionts matches those of vestimenteriferan worms more than other bivalve symbionts.

Despite the highly conserved metabolic capabilities, and the very close relationship suggested by the 16S rRNA sequences, the three phylotypes of *Thyasira* cf. *gouldi* symbiont show a large phylogenetic distance when metabolically important gene sequences are analysed. This was confirmed with multiple genes from multiple pathways. It was also found that the function of aconitate hydratase is fulfilled by the bidirectional 2-methylisocitrate dehydratase in phylotype C.

More in depth sequencing would help clarify many of the results in this paper. Binning of the reads or contigs on length, GC content, and read coverage was unsuccessful in this dataset, as all attempts resulted in one bin including the host reads. This could be due to the bacteria having a low GC content similar to the host and non-uniform read coverage. Longer reads or paired end sequencing would improve the assemblies generated and may allow for binning. Examination of individual reads could also help answer the question of population heterogeneity, however, the samples are environmental in nature and will contain bacteria that are not symbiotic. Transcriptomic, proteomic, or enzymatic tests would need to be conducted to ensure that the metabolic capabilities described here are active in the symbiont, and if so, under which conditions they occur. While culturing of symbiotic bacteria is very difficult, the environmental nature of these bacteria may make them easier to culture than obligate symbionts.

4.5 References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibeski AD, Pyshin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevner PA. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Bacteriology* 19:455-477.

Batstone RT, Laurich JR, Salvo F, Dufour SC. 2014. Divergent chemosymbiosis-related characters in *Thyasira* cf *gouldi* (Bivalvia: Thyasiridae). *PLOS One* 9.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology* 19:1-21.

Bright M, Bulgheresi S. 2010. A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology* 8:218-230.

Cavanaugh C, McKiness ZP, Newton ILG, Stewart F. 2006. Marine Chemosynthetic Symbioses. In: *The Prokaryotes* (eds. by Dworkin M, Falkow SI, Rosenberg E, Schleifer K-H, Stackebrandt E), pp. 475-507. Springer, New York.

Coleman SA, Minnick MF. 2001. Establishing a direct role for the Bartonella bacilliformis invasion-associated locus B (IalB) protein in human erythrocyte parasitism. *Infection and Immunity* 69:4373-4381.

Coleman SA, Minnick MF. 2003. Differential expression of the invasion-associated locus B (ialB) gene of Bartonella bacilliformis in response to environmental cues. *Microbial Pathogenesis* 34:179-186.

Dando PR, Spiro B. 1993. Varying nutritional dependence of the thyasirid bivalves *Thyasira sarsi* and *T. equalis* on chemoautotrophic symbiontic bacteria, demonstrated by isotope ratios of tissue carbon and shell carbonate. *Marine Ecology Progress Series* 92.

Dmytrenko O, Russell S, Loo WT, Fontanez KM, Liao L, Roeselers G, Charma R, Stewart F, Newton ILG, Woyke T, Wu D, Lang JM, Eisen JA, Cavanaugh C. 2014. The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. *BMC Genomics* 15:1-20.

Dubilier N, Bergin C, Lott C. 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Reviews Microbiology* 6:725-740.

Dufour SC. 2005. Gill anatomy and the evolution of symbiosis in the bivalve family Thyasiridae. *Biological Bulletin* 208:200-212.

Dufour SC, Laurich JR, Batstone RT, McCuaig B, Elliot A, Poduska KM. 2014. Magnetosome-containing bacteria living as symbionts of bivalves. *The ISME Journal* 8:2453-2462.

Duperron S, Gaudron SM, Rodrigues CF, Cunha MR, Decker C, Olu K. 2013. An overview of chemosynthetic symbioses in bivalves from the North Atlantic and Mediterranean Sea. *Biogeosciences* 10:3241-3267.

Duperron S, Lorion J, Samadi S, Gros O, Gaill F. 2009. Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function, and evolution. *Comptes Rendus Biologies* 332:298-310.

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*. *International Journal of Systematic and Evolutionary Microbiology* 65:2522-2530.

Fujiwara Y, Kato C, Masui N, Fujikura K, Kojima S. 2001. Dual symbiosis in the cold-seep thyasrid clam *Maorithyas hadalis* from the hadal zone in the Japan trench, western Pacific. *Marine Ecology Progress Series* 214:151-159.

Gardebrecht A, Markert S, Sievert SM. Felbeck H, Thurmer A, Albrecht D, Wollherr A, Kabisch J, Le Bris N, Lehmann R, Daniel R, Liesegang H, Hecker M, Schweder T. 2012. Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *The ISME Journal* 6:766-776.

Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* 17:377-386.

Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, Aoki Y, Inoue K, Teruya M, Satou K, Teruya K, Shimoji M, Tamotsu H, Hirano T, Maruyama T, Yoshida T. 2016. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. *The ISME Journal* 10:990-1001.

Isberg RR, Voorhis DL, Falkow S. 1987. Identification of invasin: a protein that allows enteric bacteria to penetrate cultured mammalian cells. *Cell* 50:769-778.

Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*.

Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M, Matsuyama K, Takishita K, Kawato M, Uematsu K, Fujiwara Y, Sato T, Kato C, Kitagawa M, Kato I, Maruyama T. 2007. Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Current Biology* 17:881-886.

Laurich JR, Batstone RT, Dufour SC. 2015. Temporal variation in chemoautotrophic symbiont abundance in the thyasirid bivalve *Thyasira* cf. *gouldi*. *Marine Biology* 162:2017-2028.

Markert S, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D, Albrecht D, Thürmer A, Daniel R, Kleiner M. 2011. Status quo in physiological proteomics of the uncultured Riftia pachyptila endosymbiont. *Proteomics* 11:3106-3117.

McCuaig B, Liboiron F, Dufour SC. 2017. The bivalve *Thyasira* cf. *gouldi* hosts chemoautotrophic symbiont populations with strain level diversity. *PeerJ* 5:e3597.

Passos FD, Meserani GdLC, Gros O. 2007. Structural and ultrastructural analysis of the gills in the bacterial-bearing species *Thyasira falklandica* (Bivalvia, Mollusca). *Zoomorphology* 126:153-162.

Petersen JM, Wentrup C, Verna C, Knittel K, Dubilier N. 2012. Origins and Evolutionary Flexibility of Chemosynthetic Symbionts From Deep-Sea Animals. *The Biological Bulletin* 223:123-137.

Reveillaud J, Anderson R, Reves-Sohn S, Cavanaugh C, Huber JA. 2018). Metagenomic investigation of vestimentiferan tubeworm endosymbionts from Mid-Cayman Rise reveals new insights into metabolism and diversity. *Microbiome* 6:19.

Robidart JC, Bench S, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, Allen EE, Felbeck H. 2008. Metabolic versitilty of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environmental Microbiology* 10:727-737.

Rodrigues CF, Duperron S. 2011. Distinct symbiont lineages in three thyasirid species (Bivalvia: Thyasiridae) from the easter Atlantic and Mediterranean Sea. *Naturwissenschaften* 98:281-287.

Roeselers G, Newton ILG. 2012. On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. *Applied and Environmental Microbiology* 94:1-10.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068-2069.
Shibata S, Yip ES, Quirke KP, Ondrey JM, Visick KL. 2012. Roles of the structural symbiosis polysaccharide (syp) genes in host colonization, biofilm formation, and polysaccharide biosynthesis in *Vibrio fischeri*. *Journal of Bacteriology* 194:6736-6747.
Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+ C-content biases. *Molecular Biology and Evolution* 9:678-687.
Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10:512-526.
Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. 2014. Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. *Environmental Microbiology* 16:3699-3713.

Chapter 5: Conclusions and Future Research Suggestions

5.1 Unexpected diversity in *Thyasira* cf. *gouldi* symbiont populations and among phylotypes

The main conclusion that can be made from this investigation is that the host-symbiont complex of *Thyasira* cf. *gouldi* and its bacterial symbiont is more complicated than early 16S rRNA based studies suggested (Batstone and Dufour, 2016; McCuaig et al., 2017). The three previously defined symbiont 16S rRNA phylotypes are also not as closely related, or clearly defined, as previous thought. Keeping in mind the 16S rRNA gene should be more conserved than other genes, we had nevertheless expected a high level of sequence similarity among symbiont phylotypes in other genes. However, we found unexpected diversity in functional genes that were also expected to be conserved (Chapters 2 and 4). In Chapter 2, we described two divergent RuBisCo sequences in phylotype A symbionts and suggest that the pattern is the result of horizontal gene transfer. Other functional genes, selected based on their functional role and presence in the genomes of the three phylotypes, also showed low percent identities (see Appendix tables B2-4). Based on the high 16S rRNA similarity across phylotypes, and the similar habitats, it was surprising to see the diversity in TCA cycles and biofilm related genes but note that all symbionts maintain similar overall metabolic capabilities (Chapter 4). It remains unclear whether the gene variants provide particular advantages to the symbiont under specific conditions. Different form II RuBisCo have different efficiencies in different environmental conditions (Badger and Bek, 2008). Further investigation would be needed to see if these form II RuBisCo forms provide an increased efficiency in the environments that the symbionts inhabit. These findings indicate that caution should be used when inferring physiological equivalency among bacteria based on the similarity 16S rRNA gene sequence and highlight the importance of genomic approaches to the study of symbioses.

Of the three phylotypes, phylotype C is the most genetically distinct of the symbionts (Chapter 4). This was the only rRNA phylotype to contain RuBisCo phylotype 1. It was also the only phylotype to utilize 2-methylisocitrate dehydratase within the TCA cycle. Phylotype C also contained the greatest number of unique genes, not found in either of the other symbiont assemblies (537). It also lacked syp genes (associated with biofilm formation) and is by far the rarest of the symbiont phylotypes (documented in 2/47 examined bivalves). Symbiont phylotype C was also only located in Neddy's Harbor which had distinct sediment characteristics. It was the shallowest sampling site, had the courses sediment grain size, and the lowest organic matter content (Batstone and Dufour, 2016). The distinct RuBisCo and TCA genes may be related to the sediment characteristics, making phylotype C unsuited to life in the deeper, fine grained, organic rich sediments of Deer Arm and South East Arm.

In Chapter 2, we presented evidence of symbiont population heterogeneity within host individuals. It is not uncommon for environmentally transmitted symbionts to show some population heterogeneity in juvenile hosts, however, as the host matures the symbiont population is usually selected to be very homogenous (Brissac et al., 2016). This selection is theorized to take place as the symbiont is internalized into the host cells, because the *Thyasira* cf. *gouldi* symbionts are hosted extracellularly this selection process is unlikely to occur. The diversity within a symbiont population can also be affected by when the symbionts are acquired. If the host is competent for a short time as a juvenile the symbiont population can become less diverse as the host selects bacteriocytes with the most efficient symbionts and destroys the others (Brissac et al., 2016). Life long competency coupled with lack of selection may explain the high symbiont diversity we have seen in the *Thyasira* cf. *gouldi*. Symbionts could be brought into contact with the gills by the foot which is used for burrowing and pedal feeding. This theory has

been supported by the identification of bacteria with intact magnetosomes, as well as magnetosomes from digested bacteria suggesting they were acquired at different times (Dufour et al., 2014).

A more diverse symbiont population could provide the host with a more flexible metabolism. By hosting bacteria that can fix carbon more efficiently in different environmental conditions it ensures the host has a food source if conditions change. Continual acquisition from an environmental population allows the host to acquire new symbionts that are best suited to the current conditions. This heterogeneity may also have affected the quality of our assemblies. It is very difficult to assemble highly similar genomes with genetic variation within them. The approaches used herein did not allow us to identify how heterogeneous the bacterial populations are within individuals.

5.2 Potential role for biofilm genes in controlling symbiont populations within a host

One of the unexpected findings of this research was the presence of biofilm genes in the symbiont genomes of phylotypes A and B (Chapter 4), as such genes have not yet been reported in chemoautotrophic symbionts. The symbionts appeared to contain the genes necessary for cell adhesion and secretion of an extracellular matrix. In thyasirid bivalves, symbionts occupy extracellular pockets, limited by the epithelial cell membrane and extensions of microvilli (Dufour, 2005). The ability of symbionts to create biofilms may improve their ability to colonize the host by creating an environment where the bacteria will be able to exclude competitors. The symbiotic bacteria contained many membrane transporters and possible antibiotic production genes for beta-lactamase. If the symbiont can produce an antibiotic it would allow them to exclude other bacteria, reducing competition within the gills. Biofilms could also be beneficial in the symbiont's sedimentary habitat providing many of the same benefits within the sediment.

5.3 Metabolic capabilities in the three *T.* cf. *gouldi* symbiont 16S rRNA phylotypes

Metabolic capabilities were highly conserved across all symbiont phylotypes, even in cases where individual genes were not conserved (Chapter 4). All symbionts appear to use both the sox and dsr cycles to oxidize sulfur compounds, providing energy for carbon fixation via the CBB cycle. The symbionts appear to be capable of mixotrophy, as all the draft genomes contained a full TCA cycle as well. The TCA cycle of phylotype C utilized a bifunctional 2-methylisocitrate dehydratase rather than the aconitate hydratase found in phylotypes A and B. Mixotrophy was further supported by membrane transporters, suggesting the symbiont could acquire input compounds from the environment. Mixotrophy has been described in environmental sediment bacteria such as *S. thiotaurini* (Flood et al., 2015), environmentally transmitted symbionts in vestimeniferan tubeworms (Li et al., 2018; Reveillaud et al., 2018), and has been used as evidence that symbionts are not vertically transmitted in *Solemya velum* (Dmytrenko *et al.* 2014). The reduced compounds needed for chemoautotrophy are patchy within the environment and the ability to live heterotrophically would benefit the symbiont outside the host if reduced compounds become scarce.

5.4 Further evidence for environmental transmission in *T.* cf. *gouldi* symbionts

In addition to exhibiting signs of metabolic versatility which would be adaptive in a variable sedimentary environment, the symbiont genomes show several additional characteristics of environmentally transmitted symbionts. There is no evidence of genomic reduction, they have mobile elements and many genes that are not commonly found within vertically transmitted symbionts. Symbiont phylotype B also had an extrachromosomal plasmid that encoded a type IV secretion system. The ability to move within the environment, in this case via flagella, and genes associated with sensing the environment were common to all three symbiont genomes. Again, these are often lost after vertical transmission is established, such as those symbionts found in

Calyptogena sp. (Kuwahara et al., 2007; Newton et al., 2007), and have been used as evidence of environmental transmission (Dmytrenko et al., 2014).

The symbionts are also well prepared to deal with external stresses. Each assembly contained two sets of hydrogen oxidation genes (*hox* and *huy* complexes). All genomes contained peroxidases and genes associated with oxygen sensing (*aer*). There are also many ABC transporters and multi-drug exporters identified in all the genomes. These could be utilized in defense against the host immune system, although these symbionts are extracellular (Reveillaud et al., 2018). All symbiont genomes also appear capable of assimilating nitrogen from the environment. A full suite of *nar, nos, nir* genes was found in each assembly.

5.5 Implications for the thyasirid host

The relationship of *Thyasira* cf. *gouldi* and its symbiont is very flexible. The symbiont population grows when particulate food is readily available and is reduced when external food sources are scarce. In this way the symbiont appears to act as a food storage system (Laurich et al., 2015). The evidence presented here, as well as microscopic evidence presented in Laurich et al, (2017) suggest that the host is competent throughout its lifetime. The host can be inoculated with closely related bacteria, but there is no clear selection of a single strain. This allows the host to continually be inoculated by symbiont strains that are currently successful in that environment. This is important because the conditions of the fjord sediment the host inhabits are patchy and volatile. Changing conditions may reduce the efficiency of carbon fixation in a particular strain of bacteria and make a previously inefficient strain superior. In such conditions it would be beneficial for the host to be able to "swap" symbiont strains. While the evidence was not conclusive it does appear that a single host can contain multiple symbiont strains. This would also be beneficial with changing conditions, allowing different symbiont strains to flourish under different conditions (Ikuta et al., 2016). This flexible system allows the host to readily adapt to

changing sediment conditions, and to survive for long periods of low external food availability allowing *Thyasira* cf. *gouldi* to colonize areas that may be too volatile to sustain animals which rely on an external food source.

5.6 Further Research

While this work provided a broad initial picture of the *Thyasira* cf. *gouldi* symbiont genomes, there is still much work to be done. Foremost is addressing the issue of multiple symbiont phylotypes or strains within a single host population. There are multiple ways this question can be addressed, 1) PCR of indicator genes (e.g. RuBisCo or 16S rRNA) followed by cloning into *E. coli* and resequencing; 2) resequencing of the metagenome with a more precise sequencing technology (the Ion Torrent PGM has been shown to have 1.5 homopolymer indel error per 100bp; Loman et al., 2012) followed by a bioinformatic investigation of individual reads; 3) resequencing with greater accuracy and depth, allowing for better binning and more complete assembly of the genomes in question.

While I am confident in our results, more information could be obtained through a higher quality draft genome. There are a number of ways to improve the quality of the draft created. The Ion Torrent PGM has been shown to produce more fragmented draft genomes than some of the more established next-generation sequencers; Loman et al., 2012), so alternative sequencing platforms could provide better results. There is also a wider range of programs designed for assembly of reads from these well-established sequencers than for the Ion Torrent PGM. While SPAdes does account for the specific miscalls that the Ion Torrent PGM is prone to making, such miscalls can still confound the assembly. Using paired-end reads or a technology that results in longer reads should improve the quality of the assembly. Improved quality would likely make it possible to bin the contigs, removing more of the environmental contamination, and making assembly from binned reads possible.

We have also hypothesised that the symbionts are acquired from a free-living population found in the surrounding sediment. To support this statement, the eggs of *Thyasira* cf. *gouldi* should be investigated for symbionts, through genetic testing and/or possibly microscopic studies. Evidence of living symbiotic bacteria within the sediment would also be very supportive of this theory, as to date only a partial 16S rRNA sequence has been identified within sediment samples. Partial sequences could come from dead bacteria shed from the dying/deceased hosts. Living bacterial populations could be identified by culturing (perhaps after an initial enrichment of magnetotactic bacteria from sediment samples; (Wolfe et al., 1987; Sakaguchi et al., 1996; Lins et al., 2003) or fluorescent in situ hybridization (FISH) experiments.

Regarding the metabolic capabilities of the symbiont, successful culturing followed by testing under a variety of environmental conditions would be ideal. However, culturing of symbionts outside the host is very challenging, even if they are environmentally transmitted, and initial attempts I made to culture the symbionts on thiosulfate basal medium adapted from Tuttle and Jannasch (1972) were not met with success. Culturing attempts were run in liquid media as described in the paper, semi-solid stab tubes and solid plates incubated at 4°C and 20°C. Stab tubes had loose or tight lids to help create different oxygen levels. The semi solid and solid media had Daptomycin and Fungizone added to prevent growth of other organisms. Inoculant was obtained from magnetic racetracks designed following Wolfe et al., (1987) which showed no growth, and from sediment samples, which showed growth of non-target bacteria. Transcriptomic investigations may provide some insight into which metabolic cycles are active within the host, as this may differ from what is active outside the host. Proteomics is also an option. Verifying the presence of important proteins in the clam gills would provide good

evidence of bacterial metabolism. More direct metabolic testing using living clams may also be a viable option, although lab conditions may skew the results.

In conclusion, while we have laid the groundwork for further research, many avenues of investigation could improve our understanding of this symbiotic relationship. In the short term, verification of symbiont environmental transmission and population heterogeneity might provide the greatest benefits.

5.7 References

Badger MR, Bek EJ. 2008. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. *Journal of experimental botany* 59:1525-1541.

Batstone RT, Dufour SC. 2016. Closely related thyasirid bivalves associate with multiple symbiont phylotypes. *Marine Ecology* 37:988-997.

Brissac T, Higuet D, Gros O, Merçot H. 2016. Unexpected structured intraspecific diversity of thioautotrophic bacterial gill endosymbionts within the Lucinidae (Mollusca: Bivalvia). *Marine Biology* 163:176.

Dmytrenko O, Russell S, Loo WT, Fontanez KM, Liao L, Roeselers G, Charma R, Stewart F, Newton ILG, Woyke T, Wu D, Lang JM, Eisen JA, Cavanaugh C. 2014. The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. *BMC Genomics* 15:1-20.

Dufour SC, Laurich JR, Batstone RT, McCuaig B, Elliot A, Poduska KM. 2014. Magnetosome-containing bacteria living as symbionts of bivalves. *The ISME Journal* 8:2453-2462.

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*. *International Journal of Systematic and Evolutionary Microbiology* 65:2522-2530.

Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, Aoki Y, Inoue K, Teruya M, Satou K, Teruya K, Shimoji M, Tamotsu H, Hirano T, Maruyama T, Yoshida T. 2016. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. *ISME Journal* 10:990-1001.

Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M, Matsuyama K, Takishita K, Kawato M, Uematsu K, Fujiwara Y, Sato T, Kato C, Kitagawa M, Kato I, Maruyama T. 2007. Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Current Biology* 17:881-886.

Laurich JR, Batstone RT, Dufour SC. 2015. Temporal variation in chemoautotrophic symbiont abundance in the thyasirid bivalve *Thyasira* cf. *gouldi*. *Marine Biology* 162:2017-2028.

Laurich JR, Dove R, Paillard C, Dufour SC. 2017. Life and death in facultative chemosymbioses: control of bacterial population dynamics in the Thyasiridae. *Symbiosis* 1-11.

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology* 30:434.

Li Y, Liles MR, Halanych KM. 2018. Endosymbiont genomes yield clues of tubeworm success. *The ISME Journal* 1.

Lins U, Freitas F, Keim CN, Barros HLd, Esquivel DMS, Farina M. 2003. Simple homemade apparatus for harvesting uncultured magnetotactic microorganisms. *Brazilian Journal of Microbiology* 34:111-116.

McCuaig B, Liboiron F, Dufour SC. 2017. The bivalve *Thyasira* cf. *gouldi* hosts chemoautotrophic symbiont populations with strain level diversity. *PeerJ* 5:e3597.

Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC, Fontanez KM, Lau E, Stewart F, Richardson PM, Barry KW, Saunders E, Detter JC, Wu D, Eisen JA, Cavanaugh C. 2007. The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* 315:998-1000.

Reveillaud J, Anderson R, Reves-Sohn S, Cavanaugh C, Huber JA. 2018. Metagenomic investigation of vestimentiferan tubeworm endosymbionts from Mid-Cayman Rise reveals new insights into metabolism and diversity. *Microbiome* 6:19.

Sakaguchi T, Tsujimura N, Matsunaga T. 1996. A novel method for isolation of magnetic bacteria without magnetic collection using magnetotaxis. *Journal of microbiological methods* 26:139-145.

Tuttle JH, Jannasch HW. 1972. Occurrence and types of *Thiobacillus*-like bacteria in the sea. *Limnology and Oceanography* 17(4):532-543.

Wolfe R, Thauer R, Pfennig N. 1987. A 'capillary racetrack'method for isolation of magnetotactic bacteria. *FEMS Microbiology Ecology* 3:31-35.

Appendices

Appendix A Supplemental Data for Chapter 2

Table A1: Individual Sampling Data

Sample ID	Date	GPS Coordinates		Sample Depth (m)	Shell Width (mm)	Host OTU	16S	RB
		N	W					
Deer Arm								
D 5 10/09	10/09	-	-	-	-	-	B	-
D 1 5/10	05/10	49˚33.236	57˚50.377	-	3	1	B	-
D 4 8/10	08/10	-	-	-	2	-	A	-
D 6 6/11	06/11	-	-	-	3.2	1	-	3
D 9 6/11	06/11	-	-	-	3.2	1	A	-
D 2 10/11	10/11	49° 33.222	057° 50.445	31	3.5	-	A	3
D 5.2 12/11	12/11	49° 33.219	057° 50.384	34.3	4	-	B	2
D 5.3 12/11	12/11	9° 33.219	057° 50.38	34.3	4.7	-	A	-
D 8.1 5/12	05/12	49° 33.236	057° 50.409	29	2.4	-	-	Mixed
D 12.2 5/12	05/12	49° 33.201	057° 50.431	36	3.2	-	-	2
D 12.3 5/12	05/12	49° 33.201	057° 50.431	36	4.2	-	A	3
D 13.1 10/12	10/12	49° 33.210	057° 50.420	34	4	-	B	-
D 13.2 10.12	10/12	49° 33.210	057° 50.420	34	5	-	-	Mixed
D 13.4 10/12	10/12	49° 33.210	057° 50.420	34	3.1	-	-	2
D 13.5 10/12	10/12	49° 33.210	057° 50.420	34	4.3	-	A	2
D 13.6 10/12	10/12	49° 33.210	057° 50.420	34	4.2	-	A	3
D 13.7 10/12	10/12	49° 33.210	057° 50.420	34	3.9	-	A	-
D 13.12 10/12	10/12	49° 33.210	057° 50.420	34	3.5	-	A	3
Neddy's Harbour								
N 4 5/10	05/10	49˚27.372	57˚53.280	20	-	1	B	2
N 5 5/10	05/10	49˚27.372	57˚53.280	20	-	1	C	-
N 3 8/10	08/10	-	-	-	-	2	C	-
N 4 8/10	08/10	-	-	-	-	1	B	-
N 5 8/10	08/10	-	-	-	2.5	1	B	2
N 7 8/10	08/10	-	-	-	2.5	1	B	2
N 8 8/10	08/10	-	-	-	3.5	1	B	2
N 16.1 4/11	04/11	49° 31.460	057° 52.230	16.7	2.2	1	B	2
N 16.2 4/11	04/11	49° 31.460	057° 52.230	16.7	2.8	1	B	-
N 14 6/11	06/11	-	-	-	2.3	1	-	2

N 15 6/11	06/11	-	-	-	3	-	A	2
N 20 6/11	06/11	-	-	-	3.1	1	B	2
N 21 6/11	06/11	-	-	-	3.2	1	B	2
N 27 6/11	06/11	-	-	-	2.7	1	-	2
N 1 10/11	10/11	-	-	-	3	1	B	-
N 2 10/11	10/11	-	-	-	3.2	-	C	1
N 3 10/11	10/11	-	-	-	2.5	2	C	1
N 11.3 5/12	05/12	49° 31.386	057° 52.186	29.8	3	-	-	2
N 7.1 10/12	10/12	49° 31.436	057° 52.234	-	2.5	-	-	2
N 7.2 10/12	10/12	49° 31.436	057° 52.234	-	2.3	-	-	2
N 9.1 10/12	10/12	49° 31.431	057° 52.275	-	2.5	-	B	2
N 17.1 10/12	10/12	49° 31.456	057° 52.246	-	2.7	-	B	2
N 19.1 10/12	10/12	49° 31.420	057° 52.240	-	2.6	-	-	2
N 19.2 10/12	10/12	49° 31.420	057° 52.240	-	2.6	-	-	2
N 19.3 10/12	10/12	49° 31.420	057° 52.240	-	2.5	-	B	2
South East Arm								
S 1 10/09	10/09	-	-	-	-	-	A	-
S 3 10/09	10/09	-	-	-	-	-	A	-
S 4 10/09	10/09	-	-	-	-	-	A	-
S 102B 5/10	05/10	49˚27.774	57˚43.493	30	-	2	A	3
S 103B 5/10	05/10	49˚27.774	57˚43.493	30	-	1	A	2
S 401B 5/10	05/10	49˚27.723	57˚43.455	20	5	1	A	-
S 501A 5/10	05/10	49˚27.720	57˚43.466	20	-	1	A	2
S 502A 5/10	05/10	49˚27.720	57˚43.466	20	-	1	A	3
S 1.1 4/11	04/11	49° 27. 748	057° 42. 773	27.1	4	1	A	3
S 1.2 4/11	04/11	49° 27. 748	057° 42. 773	27.1	3	1	B	2
S 1.3 4/11	04/11	49° 27.752	057° 42.449	25.3	3.5	1	-	2
S 3.1 4/11	04/11	49° 27. 751	057° 42. 822	30.2	4	1	B	2
S 3.2 4/11	04/11	49° 27. 751	057° 42. 822	30.2	3	1	A	-
S 3.3 4/11	04/11	49° 27. 751	057° 42. 822	30.2	3.5	-	A	2
S 4.2 4/11	04/11	49° 27. 745	057° 42. 806	29.9	3.5	-	A	-
S 12 6/11*	06/11	-	-	-	2.5	1	A	3
S 13 6/11*	06/11	-	-	-	4	1	A	3
S 15 6/11	06/11	-	-	-	2.7	-	A	2
S 4 10/11	10/11	-	-	-	4	-	A	2
S 6 10/11	10/11	-	-	-	3.8	-	A	-
S 8 10/11	10/11	-	-	-	3.2	1	B	2
S 13.1 12/11	12/11	49° 27.787	057° 42.482	34.6	-	1	A	-
S 13.2 12/11	12/11	49° 27.787	057° 42.482	34.6	-	-	A	3
S 15.1 12/11	12/11	-	-	-	4.2	1	B	2

S 1.1 5/12	05/12	49° 27.837	057° 42.886	31.5	5	-	-	3
S 1.4 5/12	05/12	49° 27.837	057° 42.886	31.5	33	-	-	2
S 6.4 5/12	05/12	49° 27.827	057° 42.852	30	4	-	-	3

* 16S rRNA sequences unavailable for use in phylogenies

Table A2: Variable sites within the RuBisCo sequence alignment.

	6	21	36	39	42	45	60	69	70	71	72	75	79	80	81	84	90	93	99	102	111	117
S 1.1 4/11	G	C	T	T	G	T	G	T	G	T	G	A	A	C	C	T	A	C	C	C	T	A
S 1.1 5/12	G	C	T	T	G	T	G	T	G	T	G	A	A	C	C	T	A	C	C	C	T	A
S 13 6/11	G	C	T	T	G	T	G	T	G	T	G	A	A	C	C	T	A	C	C	C	G	G
D 12.2 5/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
D 13.4 10/12	G	T	C	C	G	T	A	C	A	C	C	G	G	T	T	T	G	G	T	C	C	G
S 12 6/11	G	C	T	T	G	T	G	T	G	T	G	A	A	C	C	T	A	C	C	C	K	A
D 13.12 10/12	G	C	T	T	R	T	G	T	G	T	G	A	A	C	C	T	A	C	C	C	K	R
S 13.2 12/11	G	C	T	T	G	T	G	T	G	T	G	A	A	C	C	T	A	C	C	C	K	R
S 6.4 5/12	G	C	T	T	G	T	G	T	G	T	G	A	A	C	C	T	A	C	C	C	K	A
D 2 10/11	G	C	T	T	A	T	G	T	G	T	G	A	A	C	C	T	G	C	C	A	T	A
S 103B 5/10	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	T	G	G	T	C	C	G
D 13.5 10/12	G	T	C	C	G	Y	A	C	A	C	C	G	G	T	T	T	G	G	T	C	C	G
D 8.1 5/12	G	T	C	C	G	T	A	C	A	C	C	G	G	T	T	T	G	G	T	C	C	G
S 15 6/11	G	T	C	C	G	Y	A	C	A	C	C	G	G	Y	Y	T	G	G	T	C	C	G
S 3.3 4/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	Y	G	G	T	C	C	G
S 501A 5/10	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	Y	G	G	T	C	C	G
N 15 6/11	G	T	M	C	G	C	A	C	A	C	C	G	G	T	Y	Y	G	G	T	C	C	G
D 5.2 12/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	Y	G	G	T	C	C	G
N 16.1 4/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	Y	G	G	T	C	C	G
S 4 10/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	Y	G	G	T	C	C	G
N 19.3 10/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 21 6/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
D 13.2 10/12	K	Y	C	C	G	T	R	Y	D	B	V	R	R	Y	C	T	R	S	Y	C	G	G
D 6 6/11	G	C	T	T	G	T	G	T	G	T	G	A	R	C	C	T	A	C	C	C	Y	R
D 12.3 5/12	G	C	T	T	G	T	G	T	G	T	G	A	A	C	C	T	G	C	C	C	K	R

D 13.6 10/12	G	C	T	T	G	T	G	T	G	T	G	A	A	C	C	T	G	C	C	C	K	R
S 102B 5/10	G	C	T	T	R	T	G	T	G	T	G	A	A	C	C	T	R	C	C	M	T	A
N 11.3 5/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 14 6/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 17.1 10/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 19.1 10/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 19.2 10/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N3 10/11	G	T	T	C	G	T	C	G	A	T	G	G	G	T	T	C	C	A	T	C	T	G
N 2 10/11	G	T	T	C	G	T	C	G	A	T	G	G	G	T	T	C	C	A	T	C	T	G
N 20 6/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 27 6/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 4 5/10	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 5 8/10	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 7 8/10	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 7.1 10/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 7.2 10/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 8 8/10	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
N 9.1 10/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
S 1.2 4/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
S 1.3 4/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
S 1.4 5/12	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
S 15.1 12/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
S 3.1 4/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G
S 502A 5/10	G	C	T	T	A	T	G	T	G	T	G	A	A	C	C	T	G	C	C	A	T	A
S 8 10/11	G	T	C	C	G	C	A	C	A	C	C	G	G	T	T	C	G	G	T	C	C	G

	126	129	132	138	148	149	153	162	165	168	169	186	189	197	199	200	201	204	207	213	216	217
S 1.1 4/11	G	G	G	G	A	T	C	C	C	T	T	C	A	T	G	C	A	G	G	C	T	G
S 1.1 5/12	G	G	G	G	A	T	C	C	C	T	T	C	A	T	G	C	A	G	G	C	T	G
S 13 6/11	G	G	G	G	A	T	T	C	C	T	T	C	A	T	G	C	A	G	G	C	T	G
D 12.2 5/12	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
D 13.4 10/12	C	T	A	G	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
S 12 6/11	G	G	G	G	A	T	C	C	C	T	T	C	A	T	G	C	A	G	G	C	T	G
D 13.12 10/12	G	G	G	G	A	T	Y	C	C	T	T	C	A	T	G	C	A	G	G	C	T	G
S 13.2 12/11	G	G	G	G	A	T	Y	C	C	T	T	C	A	T	G	C	A	G	G	C	T	G
S 6.4 5/12	G	G	G	G	A	T	C	C	C	T	T	C	A	T	G	C	A	G	G	C	T	G
D 2 10/11	G	G	G	G	A	T	C	C	C	T	T	T	A	T	G	C	A	A	G	C	T	G
S 103B 5/10	C	K	A	G	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
D 13.5 10/12	C	K	A	G	G	C	C	T	T	C	C	T	M	A	A	T	C	C	A	T	C	C
D 8.1 5/12	C	K	A	G	G	C	C	T	T	C	C	T	M	A	A	T	C	C	A	T	C	C
S 15 6/11	S	K	A	G	G	C	C	T	T	C	C	T	M	W	A	T	C	C	A	T	C	C
S 3.3 4/11	C	K	A	A	G	C	C	T	T	C	C	T	M	W	A	T	C	C	A	T	C	C
S 501A 5/10	C	K	A	G	G	C	C	T	T	C	C	T	M	A	A	T	C	C	A	T	C	C
N 15 6/11	C	T	A	A	G	C	C	T	T	C	C	T	M	A	A	T	C	C	A	T	C	C
D 5.2 12/11	C	T	A	R	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 16.1 4/11	C	T	A	R	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
S 4 10/11	C	T	A	R	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 19.3 10/12	C	T	A	R	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 21 6/11	C	T	A	R	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
D 13.2 10/12	S	G	G	R	V	H	T	T	Y	H	Y	Y	A	T	A	T	C	G	R	C	Y	B
D 6 6/11	G	G	R	G	A	T	C	Y	Y	T	T	Y	A	T	R	Y	M	G	R	Y	Y	G
D 12.3 5/12	G	G	G	G	A	T	C	C	C	T	T	Y	A	T	G	C	A	R	G	C	T	G
D 13.6 10/12	G	G	G	G	A	T	C	C	C	T	T	Y	A	T	G	C	A	G	G	C	T	G
S 102B 5/10	G	G	G	G	A	T	C	C	C	T	T	Y	A	T	G	C	A	R	G	C	T	G

N 11.3 5/12	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 14 6/11	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 17.1 10/12	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 19.1 10/12	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 19.2 10/12	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N3 10/11	T	C	G	G	G	C	T	T	C	C	C	T	T	A	G	T	T	G	A	T	T	G
N 2 10/11	T	C	G	G	G	C	T	T	C	C	C	T	T	A	G	T	T	G	A	T	T	G
N 20 6/11	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 27 6/11	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 4 5/10	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 5 8/10	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 7 8/10	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 7.1 10/12	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 7.2 10/12	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 8 8/10	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
N 9.1 10/12	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
S 1.2 4/11	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
S 1.3 4/11	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
S 1.4 5/12	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
S 15.1 12/11	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
S 3.1 4/11	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C
S 502A 5/10	G	G	G	G	A	T	C	C	C	T	T	T	A	T	G	C	A	A	G	C	T	G
S 8 10/11	C	T	A	A	G	C	C	T	T	C	C	T	C	A	A	T	C	C	A	T	C	C

	218	219	231	232	234	249	252	255	258	270	276	282	285	288
S 1.1 4/11	T	C	A	C	G	C	T	T	G	C	T	G	A	T
S 1.1 5/12	T	C	A	C	G	C	T	T	G	C	T	G	A	T
S 13 6/11	T	C	A	C	G	C	T	T	G	C	T	G	A	T
D 12.2 5/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
D 13.4 10/12	C	T	G	G	T	T	C	T	C	C	C	C	C	C
S 12 6/11	T	C	A	C	G	C	T	T	G	C	T	G	A	T
D 13.12 10/12	T	C	A	C	G	C	T	T	G	C	T	G	A	T
S 13.2 12/11	T	C	A	S	G	C	T	T	G	C	T	G	A	T
S 6.4 5/12	T	C	A	C	G	C	T	T	G	C	T	G	A	T
D 2 10/11	T	C	G	C	G	C	T	C	G	T	T	G	A	T
S 103B 5/10	C	T	G	G	T	T	C	C	C	C	C	C	C	C
D 13.5 10/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
D 8.1 5/12	C	T	G	G	K	T	C	T	S	C	C	C	C	C
S 15 6/11	C	T	G	G	T	T	C	C	S	C	C	C	C	C
S 3.3 4/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
S 501A 5/10	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 15 6/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
D 5.2 12/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 16.1 4/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
S 4 10/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 19.3 10/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 21 6/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
D 13.2 10/12	Y	T	R	G	G	B	T	T	G	C	T	G	A	Y
D 6 6/11	Y	Y	R	S	G	Y	Y	T	G	C	Y	S	M	Y
D 12.3 5/12	T	C	R	C	G	C	T	Y	G	C	T	G	A	T
D 13.6 10/12	T	C	R	C	G	C	T	Y	G	C	T	G	A	T
S 102B 5/10	T	C	R	C	G	C	T	Y	G	Y	T	G	A	T
N 11.3 5/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 14 6/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 17.1 10/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 19.1 10/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 19.2 10/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N3 10/11	C	C	G	G	T	C	T	C	T	T	T	T	C	T
N 2 10/11	C	C	G	G	T	C	T	C	T	T	T	T	C	T
N 20 6/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 27 6/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 4 5/10	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 5 8/10	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 7 8/10	C	T	G	G	T	T	C	C	C	C	C	C	C	C

N 7.1 10/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 7.2 10/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 8 8/10	C	T	G	G	T	T	C	C	C	C	C	C	C	C
N 9.1 10/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
S 1.2 4/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
S 1.3 4/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
S 1.4 5/12	C	T	G	G	T	T	C	C	C	C	C	C	C	C
S 15.1 12/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
S 3.1 4/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C
S 502A 5/10	T	C	G	C	G	C	T	C	G	T	T	G	A	T
S 8 10/11	C	T	G	G	T	T	C	C	C	C	C	C	C	C

Appendix B Supplemental Data for Chapter 4

Table B1: Number of genes placed in Level 1 Subsystem functional categories by MG-RAst annotation.

level1	Phylotype A	Phylotype B	Phylotype C
Clustering-based subsystems	365	404	513
Amino Acids and Derivatives	209	234	328
Carbohydrates	175	193	362
Miscellaneous	165	179	266
DNA Metabolism	139	158	125
Protein Metabolism	176	156	306
Cell Wall and Capsule	115	137	152
Cofactors, Vitamins, Prosthetic Groups, Pigments	161	127	201
Regulation and Cell signaling	118	126	60
RNA Metabolism	118	122	148
Respiration	152	119	214
Membrane Transport	97	112	159
Stress Response	80	75	150
Virulence, Disease and Defense	48	75	81
Nitrogen Metabolism	76	65	59
Phages, Prophages, Transposable elements, Plasmids	32	64	64
Fatty Acids, Lipids, and Isoprenoids	50	45	82
Motility and Chemotaxis	63	45	80
Nucleosides and Nucleotides	48	44	118
Sulfur Metabolism	29	33	48
Phosphorus Metabolism	22	29	52

Cell Division and Cell Cycle	34	26	35
Metabolism of Aromatic Compounds	21	26	23
Potassium metabolism	18	16	26
Iron acquisition and metabolism	3	11	33
Dormancy and Sporulation	6	7	8
Secondary Metabolism	3	3	4

The phylotype C assembly was much more fragmented than the others, and may have had more duplicate genes because of this, accounting for the higher number of genes given functional annotations.

Table B2: Distance Matrix of SoxZ Genes

Base substitutions per site from between sequences are shown, all ambiguous nucleotide positions were removed from each pair. Analyses were conducted using the Kimura 2-parameter model (Kimura 1980a). A total of 315 positions were analysed in the final set using MEGA7 (Kumar *et al.* 2016).

	Phylotype A	Phylotype B
Phylotype B	0.265	
Phylotype C	0.177	0.252

Table B3: Distance Matrix of hoxS subunit beta

The number of base substitutions per site from between sequences are shown. Analyses were conducted using the Tamura 3-parameter model (Tamura 1992). The rate variation among sites was modeled with a gamma distribution (shape parameter = 5). All positions containing gaps and missing data were eliminated. There were a total of 1011 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.* 2016).

	Phylotype A	Phylotype B
Phylotype B	0.200	
Phylotype C	0.208	0.209

Table B4: Distance Matrix of dsrE

The number of base substitutions per site from between sequences are shown. Analyses were conducted using the Kimura 2-parameter model (Kimura 1980b) in MEGA7 (Kumar *et al.* 2016). The rate variation among sites was modeled with a gamma distribution (shape parameter = 5). All positions containing gaps and missing data were eliminated. There were a total of 393 positions in the final dataset.

	Phylotype A	Phylotype B
Phylotype B	0.145	
Phylotype C	0.151	0.123