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

Bonita McCuaig#, Lourdes Peña-Castilloa,b, Suzanne C. Dufoura

a Department of Biology, Memorial University of Newfoundland, St. John’s, NL, Canada

b Department of Computer Science, Memorial University of Newfoundland, St. John’s, NL, Canada
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 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 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.

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 (1-4). 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 (2, 5, 6). 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 (7, 8).

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 (6). 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 (via vertical transmission), the bacteria cannot survive without the host, and become reliant upon them (9); 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 symbioses can result in reduction or loss of the digestive tract, as nutritional reliance upon symbionts increases (6).

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 (10, 11). Some metabolic cycles of the symbionts may remove toxins present in the environment, providing the host protection from these compounds (12, 13). The sox and dsr cycles 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 (14). One approach to examining the metabolic potential of chemoautotrophic symbionts is to perform genomic, or metagenomic sequencing (15-18). 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 (19, 20). 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 (6, 21-23). Chemosymbiotic thyasirids are mixotrophs that appear to rely on particulate food to a greater extent when symbiont abundance is low (24), or at times when environmental sulfide concentrations are low (25). All thyasirid symbionts identified to date are gammaproteobacteria (23, 25, 26). The thyasirid symbionts are clustered into divergent groups which include both symbiotic and free-living sulfur-oxidizing bacteria (23). Enzymatic and PCR techniques have shown the presence of ribulosebisphosphate carboxylase (RuBisCO) and adenylylsulphate reductase in the symbionts of all chemosymbiotic thyasirids investigated (23, 25).

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 (20).
Phylogenetic analysis using 16S rRNA sequences have identified three distinct symbiont phylotypes (A – C) hosted by the two T. cf. gouldi OTUs (27, 28). 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 (27, 28). 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 (27, 28). 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 (29).

We present here the first genomic analysis of a thyasirid symbiont, that of T. cf. gouldi symbiont phylotype B (one of the most common; 27, 28). 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 transcription and protein studies.
Results and Discussion

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, accession number SAMN05358035. The assembly resulted in a highly fractured draft genome, suggesting heterogeneity in the bacterial population (recently suggested in reference 28) as genes assembled well, but intergenic spaces did not. Nevertheless, the symbiont population in the T. cf. gouldi specimen studied was 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 (16).

The assembly resulted in 12,504 contigs, with an N50 of 1500. The GC content is 42 ± 7%. In total, 20,843 putative genes were assembled and 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 1.
Table 1. The number of putative proteins assigned to level 1 subsystem functions

| 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                      |

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), the GC content (48%) is similar to that of environmental bacteria (those that have undergone genome reduction often have GC contents below 40%), and mobile elements are present (30). While none of these are conclusive evidence of environmental transmission, they are uncommon in co-evolved vertically transmitted symbionts (30); other evidence, such as the identification of the symbiont 16S rRNA gene in surrounding sediments (29), supports a free-living capability.
Four different mobile elements were identified using the MG-RAst website and PROKKA (see Methods), although the exact number of copies was unclear because of the fragmented nature of the draft genome. Phage genes were also identified, although the exact number was again indiscernible. There was no apparent loss of genes for DNA repair in the *T. cf. gouldii* symbiont genome (Table 2), in contrast to the reduced genome of the vesicomyid symbionts which lacks *recA* for genetic recombination and *mutY* for DNA repair (31). Genes for DNA repair are often lost in vertically transmitted symbionts, contributing to GC bias and the presence of many pseudogenes (30, 31).

**Table 2: Putative gene functions associated with the Calvin-Benson-Bassham Cycle**

| Putative gene function                                      | Number of copies in draft genome | Percent identity | Alignment length (bp) |
|-------------------------------------------------------------|----------------------------------|------------------|-----------------------|
| Fructose-bisphosphate aldolase class I                      | 1                                | 78.93            | 337                   |
| Fructose-bisphosphate aldolase class II                     | 1                                | 75.42            | 354                   |
| NAD-dependent glyceraldehyde-3-phosphate dehydrogenase      | 2                                | 80.39            | 225.5                 |
| NADPH-dependent glyceraldehyde-3-phosphate dehydrogenase    | 1                                | 83.54            | 328                   |
| Phosphoglycerate kinase                                     | 1                                | 71.58            | 285                   |
| Phosphoribulokinase                                         | 2                                | 80.21            | 96                    |
| Ribose 5-phosphate isomerase A                              | 4                                | 73.16            | 71.75                 |
| Ribulose bisphosphate carboxylase                           | 2                                | 83.3             | 180                   |
| Ribulose-phosphate 3-epimerase                              | 1                                | 84.68            | 222                   |
| Transketolase                                               | 1                                | 71.61            | 384                   |
| Transketolase, C-terminal section                            | 2                                | 62.81            | 112.5                 |
| Transketolase, N-terminal section                            | 1                                | 61.73            | 81                    |
| Triosephosphate isomerase                                   | 1                                | 67.07            | 167                   |

One contig contained many plasmid related genes, and further analysis uncovered the presence of a circular extrachromosomal plasmid. A *virB* operon consisting of 10 genes...
encoding a type IV secretion system was identified on the putative plasmid (Figure 1). The type IV secretion system can be used in conjunction with pili for conjugation, however, the virB operon can also be critical in both pathogenic and mutualistic relationships as a secretion system that moves molecules from bacteria to the host (32). 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 (33, 34). However, in different bacteria the system can transport a number of different molecules; notably, in some pathogenic species it can transfer small effector proteins (33). The genes virB1-5 are often transcribed together and virB7-11 form another co-transcribed group (34). On the T. cf. gouldii symbiont plasmid, the virB genes show a similar arrangement, with two hypothetical proteins placed between virB5 and virB6 (Fig. 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 the secretion system may not be important in the symbiosis, but simply used for conjugation.

The genomic data showed genes associated with flagellar assembly and function (flaG, flagA, B, C, E, F, G, H, I, J, K, L, flhA, B, F, flhD, E, G, H, K, L, M, N, P, Q, S, T, U, W, mcpB, mcpS, motB, motD, pctC, pomA, swrC, tar, ycgR, and an undefined flagellar motor protein). Also identified were the Che genes (cheA, 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 (35). 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 nutrient location and motility (29). 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 (16). Surprisingly, no magnetotaxis genes were identified by our metagenomic analysis, although magnetosome particles were identified in the symbionts of *T. cf. gouldi* (29). As magnetosome particles have been identified, associated genes are likely present in the genome, but could not be identified by the pipeline employed here.

A schematic representation of inferred metabolic capabilities of the *T. cf. gouldi* symbiont is presented in Figure 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 (36). A previous phylogenetic study (28) 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.

**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 (15). Many putative gene functions of the *T. cf. gouldi* symbiont are involved in amino acid transport and metabolism (229 assignments; Table 1), while cofactor transportation and metabolism are also frequently identified (112 times; Table 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.

**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 other bivalve chemosymbionts (37, 38). 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.

*SoxA, X, Y and Z*, which are found in a cluster in the genome of vesicomyid symbionts (37) and form a multi-enzyme system that can oxidize various forms of reduced sulfur (sulfide, thiosulfate, elemental sulfur and sulfite) to sulfate (37, 39) 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; 37). 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 20). The *T. cf. gouldi* symbiont metagenome included *soxH*, 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 (40). We also identified *cysA*, shown to import both sulfate and thiosulfate from the environment (41). Adenlylsulfate reductase subunit alpha was found, and its activity was previously detected in thyasirid symbionts (23, 25).

Many of the genes in the dsr cycle were found, with the *dsrA, B, and C* proteins as well as the peripheral *dsrE* suggesting that the pathway is running in an oxidative direction (42).
These genes as well as dsrK, 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 (16, 37, 43).

Thirteen putative functions associated with the Calvin-Benson-Bassham Cycle were discovered in the T. cf. gouldi symbiont (Table 2). The Calvin-Benson-Bassham Cycle in the T. cf gouldi symbiont utilizes a type II RuBisCO enzyme (20, 28). 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 (17, 38). 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 (15, 16). 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 (17, 38).

Hydrogen Oxidation

The symbiont also appears capable of hydrogen oxidation using the NAD⁺-reducing hydrogenase hoxHYUF, and a second set of closely related genes identified as the alpha, beta, delta, and gamma subunits of hoxS. The enzyme produced by these complexes is bi-directional. It has been described previously in the symbiont of some vestimentiferan worms (44) as well as free-living Sedimenticola selenatireducens (45).

Heterotrophy

Genes associated with the tricarboxylic acid (TCA) cycle were also identified in the T. cf. gouldi symbiont (Table 3). 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 (17). All genes for a functional TCA cycle have been found in the chemosymbiont of Solemya velum, which may occur outside of its host (17). 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 (16). 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 (36). 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.
### Table 3. Putative gene functions involved in the TCA cycle.

| Putative gene function                              | Number of copies | Percent identity | Alignment length (bp) |
|-----------------------------------------------------|------------------|------------------|-----------------------|
| 2-oxoglutarate dehydrogenase E1 component           | 1                | 70.33            | 209                   |
| Aconitate hydratase                                 | 5                | 65.53            | 109                   |
| Aconitate hydratase 2                               | 1                | 72.11            | 190                   |
| Citrate synthase (si)                               | 1                | 66.97            | 109                   |
| Dihydrolipoamide dehydrogenase                      | 2                | 73.13            | 89                    |
| Dihydrolipoamide dehydrogenase of pyruvate dehydrogenase complex | 1            | 75.83            | 211                   |
| Dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex | 2            | 75               | 138.5                 |
| Fumarate hydratase class I, aerobic                 | 2                | 88.05            | 163                   |
| Isocitrate dehydrogenase [NADP]                     | 3                | 69.21            | 265.33                |
| Malate dehydrogenase                                | 2                | 63.59            | 209                   |
| Succinate dehydrogenase flavoprotein subunit        | 4                | 77.19            | 137.5                 |
| Succinate dehydrogenase iron-sulfur protein         | 1                | 70.83            | 144                   |
| Succinyl-CoA ligase [ADP-forming] alpha chain       | 3                | 81.71            | 185                   |
| Succinyl-CoA ligase [ADP-forming] beta chain        | 4                | 72.13            | 178                   |

**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 genome. Denitrification is the process that reduces potentially harmful nitrogen compounds (nitrates, nitrites, and nitric oxide) into harmless, inert N₂ 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 protect both themselves and 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 (36). Dissimilatory nitrate respiration genes have also been identified in the symbionts of *Vesicomyosocius okutanii*, *R. pachyptila*, and *Bathymodiolus thermophilus* (16, 46, 47).

Recent work has shown some sulfur-oxidizing chemosymbionts can also fix atmospheric nitrogen into bioavailable forms (18, 48), however, we did not find any evidence of N 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 (36).

**Conclusions**

The genomic data collected on the *T. cf. gouldi* symbiont corroborates the previous data suggesting a facultative relationship, with the host clams being inoculated from the environment (28, 29). The timing of this inoculation during the host’s lifespan is still unclear and further research is needed to determine when 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. There is no evidence of genome reduction in these symbionts, and the genomic 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 (29).

The metabolic capabilities of the symbionts are comparable to previously described sulfur oxidizing bacteria. The symbionts utilize multiple pathways for sulfur oxidation, both sox and dsr, and the Calvin-Benson-Bassham 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 (36).

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.

**Methods**

**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. 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. Total DNA was extracted from the gills of a single individual (host OTU 1; reference 20) 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 Torrent Fragmentation Kit 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.

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 (49, 50); 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 binned using MEGAN5 (51).

Assembly of the binned data was conducted using SPAdes (52). Ion Torrent specific settings with kmers 27, 35, 55 and 77 were used. All bacterial, unknown, and unassigned reads were used for assembly. Contigs of 200 bp or more were entered into the pipeline. Annotation was run using the MG-RAST website (http://metagenomics.anl.gov/) (53), 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 (54).

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.
References

1. Stewart FJ, Newton ILG, Cavanaugh C. Chemosynthetic endosymbioses: adaptations to oxic-anoxic interfaces. TRENDS in Microbiology. 2005;13(9):439-48.

2. Cavanaugh C, McKiness ZP, Newton ILG, Stewart F. Marine Chemosynthetic Symbioses. In: Dworkin M, Falkow SI, Rosenberg E, Stackebrandt E, editors. The Prokaryotes. New York: Springer; 2006. p. 475-507.

3. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nature Reviews Microbiology. 2008;6:725-40.

4. Moya A, Pereto J, Gil R, Latorre A. Learning to live together: genomic insights into prokaryote-animal symbioses. Nature Reviews Genetics. 2008;9:218-29.

5. Passos FD, Meserani GdLC, Gros O. Structural and ultrastructural analysis of the gills in the bacterial-bearing species *Thyasira falklandica* (Bivalvia, Mollusca). Zoomorphology. 2007;126:153-62.

6. Roeselers G, Newton ILG. On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. Applied and Environmental Microbiology. 2012;94:1-10.

7. Gündüz EA, Douglas A. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. Proceedings of the Royal Society of London B: Biological Sciences. 2009;276(1658):987-91.

8. Stanley SM. Evolutionary radiation of shallow-water Lucinidae (Bivalvia with endosymbionts) as a result of the rise of seagrasses and mangroves. Geology. 2014;42(9):803-6.

9. Bennett GM, Moran NA. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proceedings of the National Academy of Sciences. 2015;112(33):10169-76.

10. Felbeck H, Jarchow J. Carbon Release from Purified Chemoautotrophic Bacterial Symbionts of the Hydrothermal Vent Tubeworm *Riftia pachyptila*. Physiological Zoology. 1998;71(3):294-302.

11. Bright M, Keckeis H, Fisher CR. An autoradiographic examination of carbon fixation, transfer and utilization in the *Riftia pachyptila* symbiosis. Marine Biology. 2000;136:621-32.
12. Waite TJ, Moore TS, Childress JJ, Hsu-Kim H, Mullaugh KM, Nuzzio DB, et al. Variation in sulfur speciation with shellfish presence at a Lau Basin diffuse flow vent site. Journal of Shellfish Research. 2008;27(1):163-8.

13. Podowski EL, Ma S, Luther III GW, Wardrop D, Fisher CR. 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. 2010;418:25-45.

14. Liao L, Wankel SD, Wu M, Cavanaugh CM, Girguis PR. Characterizing the plasticity of nitrogen metabolism by the host and symbionts of the hydrothermal vent chemoautotrophic symbioses *Ridgeia piscesae*. Molecular ecology. 2014;23(6):1544-57.

15. Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC, et al. The *Calyptogena magnifica* chemoautotrophic symbiont genome. Science. 2007;315:998-1000.

16. Robidart JC, Bench S, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, et al. Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. Environmental Microbiology. 2008;10(3):727-37.

17. Dmytrenko O, Russell S, Loo WT, Fontanez KM, Liao L, Roeselers G, et al. The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. BMC Genomics. 2014;15(1):1-20.

18. König S, Gros O, Heiden SE, Hinzke T, Thuerner A, Poehlein A, et al. Nitrogen fixation in a chemoautotrophic lucinid symbiosis. Nature microbiology. 2017;2(1):16193.

19. Taylor JD, Williams ST, Glover EA. Evolutionary relationships of the bivalve family Thyasiridae (Mollusca: Bivalvia), monphyly and superfamily status. Journal of the Marine Biological Association of the United Kingdom. 2007;87:565-74.

20. Batstone RT, Laurich JR, Salvo F, Dufour SC. Divergent chemosymbiosis-related characters in *Thyasira cf gouldi* (Bivalvia: Thyasiridae). PLOS One. 2014;9(3).
21. Dufour SC. Gill anatomy and the evolution of symbiosis in the bivalve family Thyasiridae. Biological Bulletin. 2005;208:200-12.

22. Taylor JD, Glover EA. Chemosymbiotic Bivalves. In: Kiel S, editor. The Vent and Seep Biota. Topics and Geobiography. 33: Springer Science+Business Media B.V.; 2010.

23. Rodrigues CF, Duperron S. Distinct symbiont lineages in three thyasirid species (Bivalvia: Thyasiridae) from the easter Atlantic and Mediterranean Sea. Naturwissenschaften. 2011;98:281-7.

24. Dufour SC, Felbeck H. Symbiont abundance in thyasirids (Bivalvia) is related to particulate food and sulphide availability. Marine Ecology Progress Series. 2006;320:185-94.

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

26. Fujiwara Y, Kato C, Masui N, Fujikura K, Kojima S. Dual symbiosis in the cold-seep thyasrid clam Maorithyas hadalis from the hadal zone in the Japan trench, western Pacific. Marine Ecology Progress Series. 2001;214:151-9.

27. Batstone RT, Dufour SC. Closely related thyasirid bivalves associate with multiple symbiont phylotypes. Marine Ecology. 2016;37(5):988-97.

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

29. Dufour SC, Laurich JR, Batstone RT, McCuaig B, Elliot A, Poduska KM. Magnetosome-containing bacteria living as symbionts of bivalves. The ISME Journal. 2014;8:2453-62.

30. Moran NM, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. Annual Review of Genetics. 2008;42:165-90.
31. Kuwahara H, Takaki Y, Shimamura S, Yoshida T, Maeda T, Kunieda T, et al. Loss of genes for DNA recombination and repair in the reductive genome evolution of thioautotrophic symbionts of *Calyptogena* clams. BMC Evolutionary Biology. 2011;11(285).

32. Dale C, Moran NM. Molecular interactions between bacterial symbionts and their hosts. Cell. 2006;126:453-65.

33. Christie PJ, Vogel JP. Bacterial type IV secretion: conjugative systems adapted to deliver effector molecules to host cells. TRENDS in Microbiology. 2000;8(8):354-60.

34. Christie PJ. Type IV secretion: *Agrobacterium* VirB/D4 and related conjugation systems. Biochimica et Biophysica. 2004:219-34.

35. Ferrández A, Hawkins AC, Summerfield DT, Harwood CS. Cluster II che genes from *Pseudomonas aeruginosa* are required for an optimal chemotactic response. Journal of bacteriology. 2002;184(16):4374-83.

36. Flood BE, Jones DS, Bailey JV. *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. 2015;65:2522-30.

37. Harada M, Yoshida T, Kuwahara H, Shimamura S, Takaki Y, Kato C, et al. Expression of genes for sulfur oxidation in the intracellular chemoautotrophic symbiont of the deep-sea bivalve *Calyptogena okutanii*. Extremophiles. 2009;13:895-903.

38. Roeselers G, Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, et al. Complete genome sequence of *Candidatus Ruthia magnifica*. Standards in Genomic Sciences. 2010;3:163-73.

39. Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J. Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? Applied and environmental microbiology. 2001;67(7):2873-82.
40. Rother D, Henrich H-J, Quentmeier A, Bardischewsky F, Friedrich CG. 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. 2001;183(15):4499-508.

41. Sirko A, Hryniewicz M, Hulanicka D, Bock A. Sulfate and thiosulfate transport in *Escherichia coli* K-12: Nucleotide sequence and expression the *cysTWAM* gene cluster. Journal of Bacteriology. 1990;172(6):3351-7.

42. Bradley AS, Leavitt WD, Johnston DT. Revisiting the dissimilatory sulfate reduction pathway. Geobiology. 2011;9:446-57.

43. Nakagawa S, Shimamura S, Takaki Y, Suzuki Y, Murakami S-i, Watanabe T, et al. Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. The ISME Journal. 2014;8(1):40-51.

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

45. Narasingarao P, Häggblom MM. *Sedimenticola selenatireducens*, gen. nov., sp. nov., an anaerobic selenate-respiring bacterium isolated from estuarine sediment. Systematic and applied microbiology. 2006;29(5):382-8.

46. Kleiner M, Petersen JM, Dubilier N. Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. Current Opinion in Microbiology. 2012;15:621-31.

47. Ponnuudurai R, Sayavedra L, Kleiner M, Heiden SE, Thürmer A, Felbeck H, et al. Genome sequence of the sulfur-oxidizing *Bathymodiolus thermophilus* gill endosymbiont. Standards in genomic sciences. 2017;12(1):50.
48. Petersen JM, Kemper A, Gruber-Vodicka H, Cardini U, van der Geest M, Kleiner M, et al. Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. Nature microbiology. 2016;2:16195.

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

50. Afgan E, Baker D, Van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic acids research. 2016;44(W1):W3-W10.

51. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome research. 2007;17(3):377-86.

52. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Bacteriology. 2012;19(5):455-77.

53. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008;9(386).

54. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068-9.
