McCuaig et al. Animal Microbiome (2020) 2:7
https://doi.org/10.1186/s42523-020-00025-9

Animal Microbiome

RESEARCH ARTICLE Open Access

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

Bonita McCuaig[1], Lourdes Peña-Castillo[1,2] and Suzanne C. Dufour[1*]

Abstract

Background: Next-generation sequencing has opened new avenues for studying metabolic capabilities of bacteria that cannot be cultured. Here, we provide a metagenomic description of chemoautotrophic gammaproteobacterial symbionts associated with *Thyasira* cf. *gouldi*, a sediment-dwelling bivalve from the family Thyasiridae. Thyasirid symbionts differ from those of other bivalves by being extracellular, and recent work suggests that they are capable of living freely in the environment.

Results: *Thyasira* cf. *gouldi* symbionts appear to form mixed, non-clonal populations in the host, show no signs of genomic reduction and contain many genes that would only be useful outside the host, including flagellar and chemotaxis genes. The thyasirid symbionts may be capable of sulfur oxidation via both the sulfur oxidation and 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.

Conclusions: In this study, we have generated the first thyasirid symbiont genomic resources. In *Thyasira* cf. *gouldi*, symbiont populations appear non-clonal and encode genes for a plethora of metabolic capabilities; future work should examine whether symbiont heterogeneity and metabolic breadth, which have been shown in some intracellular chemosymbionts, are signatures of extracellular chemosymbionts in bivalves.

Keywords: Sulfur-oxidizing, Thyasiridae, Magnetotactic, Facultative, Mollusc, Mixotrophic

Background

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. In horizontal transmission, hosts are inoculated by symbionts released by nearby adults, whereas in environmental transmission, juveniles are inoculated from a free-living symbiont population [3, 6]. While the latter mode of transmission does not guarantee symbiont transfer, it does confer some advantages to both partners. Bacteria that maintain

* Correspondence: sdufour@mun.ca
[1]Department of Biology, Memorial University of Newfoundland, St. John's, NL, Canada
Full list of author information is available at the end of the article

© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

a free-living population can avoid genomic reduction, which, in some vertically transmitted bacteria, can result in the deletion of genes that no longer improve their fitness. Once genomic reduction occurs (via vertical transmission), some symbionts may not be able to survive without their host, and become reliant upon them [9]; this does not occur in symbionts that maintain a functional environmental population. The maintenance of variation within free-living 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 present in their progenitors. Environmentally acquired symbionts could potentially form diverse or heterogeneous populations within host individuals if multiple symbionts are acquired from a genetically diverse founding population.

Symbionts 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]. In chemoautotrophic symbionts, the sulfur oxidation (sox) and dissimilatory sulfate reduction (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 [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, the metabolic capabilities of symbionts can be inferred.

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]. The symbionts are found in high abundance among the elongate microvilli of abfrontally expanded gill filaments [20]. This relationship has been observed over multiple years in three sampling sites, and all specimens of *T.* cf. *gouldi* OTUs 1 and 2 whose gills have been examined using thin sectioning and transmission electron microscopy (i.e., over 200 specimens) harboured large populations of visibly homogenous bacterial symbionts [20, 24, 27, 28]. Phylogenetic analysis using 16S rRNA gene sequences have identified three distinct symbiont phylotypes (A – C) hosted by the two symbiotic *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 evidence of multiple RuBisCO types in a single host [27, 28]. The three bacterial 16S rRNA gene 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]. Stable isotope analysis of *T.* cf. *gouldi* supports chemoautotrophic activity and nutrient transfer from bacteria to host, notably through a lower $\delta^{15}N$ value than in non-symbiotic, co-occurring bivalves (representing a greater contribution of nutrients from a low trophic level, likely through chemoautotrophic primary production); tissue $\delta^{13}C$ values were less negative than in other chemosymbiotic thyasirids, as expected due to the presence of RuBisCO type II [29]. Taken together, the consistent presence of abundant gill-associated sulfur-oxidizing bacteria in numerous *T.* cf. *gouldi* individuals and the stable isotope data are strongly indicative of a symbiotic relationship rather than an environmental contamination. *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 [30].

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. Based on 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

Overview of metagenome

Our analysis recovered a large number of contigs (12, 504, with an N_{50} of 1870) that we consider to represent the symbiont population's metagenome; these contigs could not be assembled into a single genome. A high level of sequence redundancy, with sequences of 54% of contigs being significantly similar to at least one other contig (Blast Evalue $< 10^{-10}$; Additional file 1) suggests many cases of slight sequence divergence across the symbiont (population) metagenome. The contigs from our metagenome show highly similar trinucleotide and tetranucleotide frequency distributions (i.e. tight clusters in Principal Components Analyses; Additional file 1), supporting our interpretations that: 1) the sequences we retained as belonging to symbionts (i.e., the metagenome) come from highly similar bacteria; and 2) the metagenome contained few contaminants, as expected given the very high abundance of morphologically similar symbionts observed on *Thyasira* cf. *gouldi* gills using transmission electron microscopy [20]. This metagenome is most likely a combination of multiple bacterial genomes recovered from a population of symbionts that share the same 16S rRNA gene sequence (only one complete 16S rRNA gene sequence was found, for phylotype B), but with divergence in gene arrangement and gene sequence across the population. Most likely, the genomes across the symbiont population were highly similar overall, but contained regions of sufficient variation to prevent assembly of contigs into a single genome, accounting for our inability to quantify genomic size. The heterogeneous (non-clonal) nature of thyasirid symbiont populations was suggested previously [28]. Hereafter, we describe features of this metagenome, considering that it represents a similar, yet divergent population of *T.* cf. *gouldi* symbionts sharing the same 16S rRNA phylotype B gene sequence.

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%, lower than the symbionts of *Solemya velum* (51%) and *Riftia pachyptila* (58%) and higher than the symbionts of *Calyptogena magnifica* (34%) and *Calyptogena okutanii* (32%); it is similar to the GC content of the free-living bacteria *Thiomicrospira crunogena* (43%) [17]. In total, 20,843 putative open reading frames (ORFs) were detected in the contigs: this number likely includes variants of the same genes present within different genomes across the metagenome. Using the MG-RAST workflow, possible functions were assigned to 3339 of the identified ORFs, allowing us to infer some of the metabolic capabilities of the *T.* cf. *gouldi* symbionts. Assigned gene functions were grouped into subsystems (curated collections of genes associated with specific metabolic pathways, biological processes, structural complexes or protein families [31]) using the MG-RAST workflow (Table 1).

Genomic support of environmental transmission

Some general characteristics of the metagenome are consistent with *T.* cf. *gouldi* symbionts having a free-living stage: there is no sign of genomic reduction (genes present were assigned to a broad range of functions, with no essential category appearing to be missing) and mobile elements are present [32]. While this is not conclusive evidence of environmental transmission, they are uncommon in co-evolved vertically transmitted

Table 1 The number of genes assigned to functional subsystems using MG-RAST

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

symbionts [32]; other evidence, such as the identification of the symbiont 16S rRNA gene in surrounding sediments [30], supports a free-living capability.

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 because of the fragmented nature of metagenome. 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 metagenome (Table 2), in contrast to the reduced genome of the vesicomyid symbionts which lack *recA* for genetic recombination and *mutY* for DNA repair [33]. The loss of genes for DNA repair has been observed in vertically transmitted symbionts, contributing to GC bias and the presence of many pseudogenes [32, 33].

The metagenomic data showed genes associated with flagellar assembly and function (*flaG*, *flgA*, *B*, *C*, *E*, *F*, *G*, *H*, *I*, *J*, *K*, *L*, *flhA*, *B*, *F*, *fliD*, *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 [34]. These genes are essential to locate and move to the micro-oxic, 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 micro-oxic areas where sulfur oxidation is most efficiently carried out [35]. 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 most likely require key genes for substrate location and motility [30]. 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* [30]. Magnetosome genes are likely present in the genome, but either were not sequenced or could not be identified by the pipeline employed here. A directed BLAST search using known magnetosome genes *mam*, *man*, *mms*, and *mad* did not yield any results. Interestingly, these genes have not yet been identified in magnetotactic gammaproteobacteria; this taxon may simply lack comparable genetic information.

A schematic representation of inferred metabolic capabilities of the *T.* cf. *gouldi* symbiont is presented in Fig. 1. 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 oxic conditions, where it must instead rely on heterotrophy [36]. A previous phylogenetic study [28] placed the *T.* cf. *gouldi* symbiont 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 we cannot validate this theory.

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

Putative gene function	Number of copies in the metagenome	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

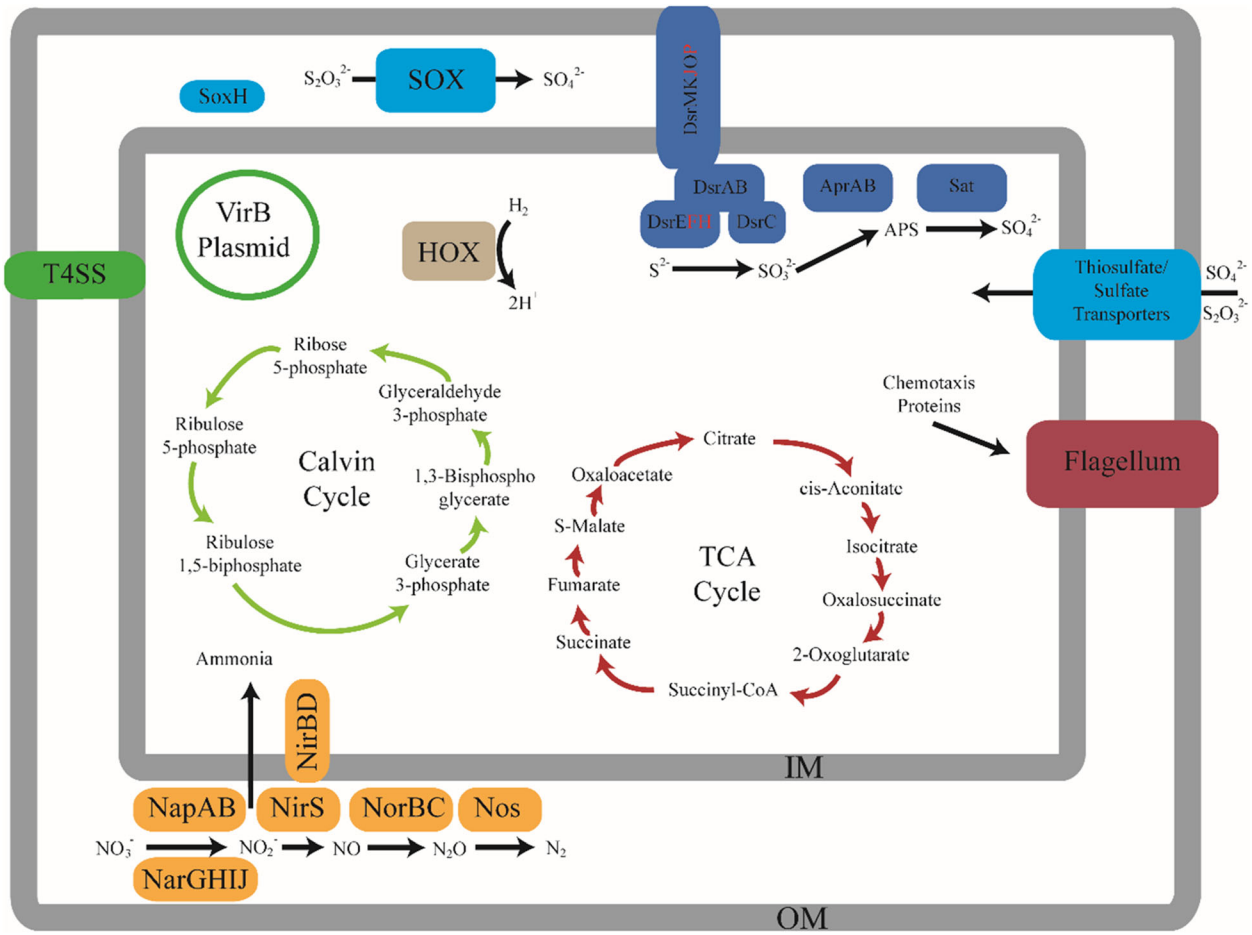

Fig. 1 Schematic representation of inferred metabolic capabilities of the *Thyasira* cf. *gouldi* symbiont. The major metabolic cycles discussed in this paper are highlighted within the figure. The inner membrane is designated by IM and the outer membrane by OM. Genes that were not identified within either annotation pipeline are designated by a red letter

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 transport 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 (likely an oxidative dsr pathway) 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 [37–39]. 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 multienzyme system that can oxidize various forms of reduced sulfur (sulfide, thiosulfate, elemental sulfur and sulfite) to sulfate [37, 40] 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 [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 [41]. We also identified *cysA*, shown to import both sulfate and thiosulfate from the environment [42]. Adenylylsulfate reductase 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 [43]. These genes as well as *dsrK*, *M*, *R*, *S* are present in the symbiont genome. The vestimentiferan symbionts discussed by Li et al. did not contain the *dsrMKJOP* membrane bound protein complex [39]. 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, 44].

Thirteen putative functions associated with the Calvin-Benson-Bassham Cycle were discovered in the *T.* cf. *gouldi* symbiont metagenome (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 [39, 45] as well as free-living *Sedimenticola selenatireducens* [46].

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 seasonally variable environments where sulfur compounds can be scarce, and availability of organic matter fluctuates.

Table 3 Putative gene functions involved in a complete TCA cycle identified by MG-RAST

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 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 [36]. Dissimilatory nitrate respiration genes have also been identified in the symbionts of *Vesicomyosocius okutanii*, *Bathymodiolus thermophilus*, and a number of vestimentiferan tubeworms [16, 39, 47, 48].

Recent work has shown some sulfur-oxidizing chemosymbionts can also fix atmospheric nitrogen into bioavailable forms [18, 49], 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 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 [28, 30]. 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 [30].

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 ATP generation 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. Future studies would benefit from the application of more powerful sequencing technologies, such as mate-paired sequencing or single-cell genomic sequencing, which would facilitate genomic comparison with other symbionts and free living bacteria.

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. Host individuals were collected during times of high symbiont abundance [50], and visually inspected for indication of high symbiont population. To reduce environmental contamination of bacteria not contained within the mucus between gill filaments, surface of the gill was rinsed before DNA extraction. Total DNA was extracted from the gills of a single individual (host OTU 1 [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 on the Galaxy Website (usegalaxy.org) and FastQC software [51, 52]; 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 into contigs and then binned using MEGAN5 [53], which used BLAST to compare each contig to the nr database. Reads from all contigs identified as "bacteria", "not assigned", "unclassified" or "no hits" were identified and retained for reassembly and subsequent analysis; the combination of these reads was deemed to represent the symbiont metagenome.

Assembly of the binned data (i.e., deemed to represent the symbiont metagenome) was conducted using SPAdes [54]. Ion Torrent specific settings with kmers 27, 35, 55 and 77 were used. Annotation was run using the MG-RAST website (http://metagenomics.anl.gov/) [55], 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 [56].

Principal component analysis (PCA) was performed on the trinucleotide and tetranucleotide frequency distribution of the read sequences. All 5.5 million reads were mapped back to the contigs assembled by SPAdes using BWA-MEM [57] version 0.7.17 with default settings. The 25% of reads that did not align to the contigs were identified using SAMtools [58]. Trinucleotide and tetranucleotide frequency distribution of the read sequences were obtained using the R function oligonucleotideFrequency available in the Biostrings package [59]. PCA was carried out using the R function prcomp with scale and center set to true. PCA results were visualized using the R function autoplot available in the ggfortify package [60]. Reads were coloured based on whether they were or not aligned to the contigs. To investigate whether the contigs identified by MEGAN5 as "eukaryote" have a distinct trinucleotide frequency distribution from the contigs deemed belonging to the bacterial metagenome, PCA as described above was performed on the trinucleotide frequency distribution of the contigs sequences.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s42523-020-00025-9.

Additional file 1: Figure S1. Total number of contigs (log2 scale) with a significant (Evalue <10e-10) local sequence similarity with *N* other contigs. *N* is indicated on the x-axis. Approximately 55% of the 12,504 contigs assembled have significant local sequence similarity with at least one other contig. **Figure S2.** PCA of the trinucleotide (left) and tetranucleotide (right) frequency distribution of the raw reads. Aligned "Yes" indicates the reads that were aligned back to the contigs deemed to belong to the symbiont metagenome. On each axis label the number between brackets is the percentage of variance explained by the corresponding principal component. Ellipses are 95% data ellipses assuming a multivariate Gaussian distribution. Ellipses have the mean vector as their center and cover 95% of the corresponding data points. The reads aligned to the metagenome have low variance in terms of their tri(tetra)nucleotide frequency. **Figure S3.** PCA of the trinucleotide frequency distribution of the putative symbiont contigs (metagenome) and eukaryotic contigs as classified by MEGAN5. On each axis label the number between brackets is the percentage of variance explained by the corresponding principal component. Ellipses are 95% data ellipses assuming a multivariate Gaussian distribution. Ellipses have the mean vector as their center and cover 95% of the corresponding data points. The symbiont metagenome have low variance in terms of their trinucleotide frequency distribution and are clearly more homogenous than the eukaryotic contigs.

Abbreviations

ATP: Adenosine triphosphate; C: Carbon; dsr: Dissimilatory sulfate reduction; IM: Inner membrane; N: Nitrogen; NAD: Nicotinamide adenine dinucleotide; NADPH: Dihydronicotinamide-adenine dinucleotide phosphate; nir: Nitrate reduction; OM: Outer membrane; ORF: Open reading frame; OTU: Operational taxonomic unit; PCA: Principal components analysis; PCR: Polymerase chain reaction; PTS: Phosphotransferase system; rRNA: Ribosomal RNA; RuBisCO: Ribulose-1,5-bisphosphate carboxylase/oxygenase; sox: Sulfur oxidation; TCA: Tricarboxylic acid

Acknowledgements

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

Authors' contributions

BM performed sampling, sequencing and bioinformatics analyses, and wrote the manuscript. LPC assisted with bioinformatics analyses and provided editorial assistance. SCD participated in sampling and data interpretation, and contributed to writing the manuscript. All authors read and approved the final manuscript.

Funding

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. NSERC played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

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.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details
[1]Department of Biology, Memorial University of Newfoundland, St. John's, NL, Canada. [2]Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada.

Received: 4 November 2019 Accepted: 20 February 2020
Published online: 05 March 2020

References

1. Stewart FJ, Newton ILG, Cavanaugh C. Chemosynthetic endosymbioses: adaptations to oxic-anoxic interfaces. Trends Microbiol. 2005;13:439–48.
2. Cavanaugh C, McKiness ZP, Newton ILG, Stewart F. Marine chemosynthetic symbioses. In: Dworkin M, Falkow SI, Rosenberg E, Schleifer K-H, 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. Nat Rev Microbiol. 2008;6:725–40.
4. Moya A, Pereto J, Gil R, Latorre A. Learning to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet. 2008;9:218–29.
5. Passos FD, Meserani GLC. 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. Appl Environ Microb. 2012; 94:1–10.
7. Gündüz EA, Douglas A. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. P Roy Soc B-Biol Sci. 2009;276:987–91.
8. Stanley SM. Evolutionary radiation of shallow-water Lucinidae (Bivalvia with endosymbionts) as a result of the rise of seagrasses and mangroves. Geology. 2014;42:803–6.
9. Bennett GM, Moran NA. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. P Natl Acad Sci USA. 2015;112:10169–76.
10. Felbeck H, Jarchow J. Carbon release from purified chemoautotrophic bacterial symbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. Physiol Zool. 1998;71:294–302.
11. Bright M, Keckeis H, Fisher CR. An autoradiographic examination of carbon fixation, transfer and utilization in the *Riftia pachyptila* symbiosis. Mar Biol. 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. J Shellfish Res. 2008;27:163–8.
13. Podowski EL, Ma S, Luther GW III, 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. Mar Ecol Prog Ser. 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*. Mol Ecol. 2014;23:1544–57.
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. Environ Microbiol. 2008;10:727–37.
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;15:1–20.
18. König S, Gros O, Heiden SE, Hinzke T, Thuermer A, Poehlein A, et al. Nitrogen fixation in a chemoautotrophic lucinid symbiosis. Nat Microbiol. 2017;2:16193.
19. Taylor JD, Williams ST, Glover EA. Evolutionary relationships of the bivalve family Thyasiridae (Mollusca: Bivalvia), monophyly and superfamily status. J Mar Biol Assoc UK. 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:e92856.
21. Dufour SC. Gill anatomy and the evolution of symbiosis in the bivalve family Thyasiridae. Biol Bull. 2005;208:200–12.
22. Taylor JD, Glover EA. Chemosymbiotic Bivalves. In: Kiel S, editor. The Vent and Seep Biota. Topics and Geobiology. 33: Springer Science+Business Media B.V.; 2010. pp. 107–135.
23. Rodrigues CF, Duperron S. Distinct symbiont lineages in three thyasirid species (Bivalvia: Thyasiridae) from the eastern 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. Mar Ecol Prog Ser. 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. Mar Ecol Prog Ser. 1993;92:151–8.
26. Fujiwara Y, Kato C, Masui N, Fujikura K, Kojima S. Dual symbiosis in the cold-seep thyasirid clam *Maorithyas hadalis* from the hadal zone in the Japan trench, western Pacific. Mar Ecol Prog Ser. 2001;214:151–9.
27. Batstone RT, Dufour SC. Closely related thyasirid bivalves associate with multiple symbiont phylotypes. Mar Ecol. 2016;37:988–97.
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. Zanzerl H, Salvo F, Jones SW, Dufour SC. Feeding strategies in symbiotic and asymbiotic thyasirid bivalves. J Sea Res. 2019;145:16–23.
30. Dufour SC, Laurich JR, Batstone RT, McCuaig B, Elliot A, Poduska KM. Magnetosome-containing bacteria living as symbionts of bivalves. ISME J. 2014;8:2453–62.
31. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 2005;33:5691–702.
32. Moran NM, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. 2008;42:165–90.
33. 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 Evol Biol. 2011;11:285.
34. Ferrández A, Hawkins AC, Summerfield DT, Harwood CS. Cluster II che genes from *Pseudomonas aeruginosa* are required for an optimal chemotactic response. J Bacteriol. 2002;184:4374–83.
35. Brune A, Frenzel P, Cypionka H. Life at the oxic-anoxic interface: microbial activities and adaptations. FEMS Microbiol Rev. 2000;24:691–710.
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*. Int J Syst Evol Micr. 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. Stand Genomic Sci. 2010;3:163–73.
39. Li Y, Liles MR, Halanych KM. Endosymbiont genomes yield clues of tubeworm success. ISME J. 2018;12:2785–95.
40. Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J. Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? Appl Environ Microbiol. 2001;67:2873–82.
41. 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. J Bacteriol. 2001;183:4499–508.
42. 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. J Bacteriol. 1990;172:3351–7.
43. Bradley AS, Leavitt WD, Johnston DT. Revisiting the dissimilatory sulfate reduction pathway. Geobiology. 2011;9:446–57.
44. Nakagawa S, Shimamura S, Takaki Y, Suzuki Y, Murakami S, Watanabe T, et al. Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. ISME J. 2014;8:40–51.
45. 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;6:19.
46. Narasingarao P, Häggblom MM. *Sedimenticola selenatireducens*, gen. Nov., sp. nov., an anaerobic selenate-respiring bacterium isolated from estuarine sediment. Syst Appl Microbiol. 2006;29:382–8.
47. Kleiner M, Petersen JM, Dubilier N. Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. Curr Opin Microbiol. 2012;15:621–31.

48. Ponnudurai R, Sayavedra L, Kleiner M, Heiden SE, Thürmer A, Felbeck H, et al. Genome sequence of the sulfur-oxidizing *Bathymodiolus thermophilus* gill endosymbiont. Stand Genomic Sci. 2017;12:50.
49. 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. Nat Microbiol. 2016;2:16195.
50. Laurich JR, Batstone RT, Dufour SC. Temporal variation in chemoautotrophic symbiont abundance in the thyasirid bivalve *Thyasira* cf. *gouldi*. Mar Biol. 2015;162:2017–28.
51. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, et al. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol. 2010;19(10):1–21.
52. 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 Res. 2016;44:W3–W10.
53. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17:377–86.
54. 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. J Bacteriol. 2012;19:455–77.
55. 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.
56. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
57. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv e-prints. 2013; arXiv:1303.3997. (pre-print).
58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G. Durbin R and 1000 genome project data processing subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25: 2078–9.
59. Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. R package version 2.52.0; 2019.
60. Tang Y, Horikoshi M, Li W. Ggfortify: unified interface to visualize statistical results of popular R packages. R J. 2016;8:474–85.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC