Pearl in the Mud: Genome Assembly and Binning of a cold seep *Thiomargarita nelsonii* cell and Associated Epibionts from an Environmental Metagenome

A THESIS
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Palmer Scott Fliss

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Adviser Jake V. Bailey

January 2014

Acknowledgements

I would like to thank Jake Bailey for his guidance and mentorship throughout my time at the University of Minnesota. I would also like to profusely thank Beverly Flood her guidance, support, and teachings on microbiology, geobiology, and acceptance into what was once and remains a challenging field, in addition to her extensive work on the DNA extraction and wet lab microbiology. Dan Jones was an invaluable asset to my understanding of science of bioinformatics and geobiology, his talks and insights were key to understanding the breadth of this project. The assistance of the aforementioned scientists was critical to the success of this project. I would like to extend my gratitude to the entire Bailey lab group for their discussions, friendship, knowledge, and aid both scholarly and emotional. Many thanks to Greg Dick and Sunit Jain for their computational resources and their continued support and suggestions on best practices regarding metagenomic assembly and binning. Thanks to Greg Rouse, Victoria Orphan, and Lisa Levin for assistance with the collection of gastropods from Hydrate Ridge. Additional thanks to Ying Zhang, and the MSI Bioinformatics support staff for their assistance in troubleshooting assemblies and software installation. Charles Nguyen and the rest of the Earth Sciences IT Department were essential to the success of establishing the Bailey Geobiology Lab Server, and their assistance with installation and structuring our data was profoundly helpful. Additional thanks to Anne-Kristin Kaster for generously donating PacBio RS sequencing wells, as well as her advice regarding assembly and data synthesis. I would like to thank D. Fox and D. Knights for serving on my thesis defense committee. Portions of this work were supported by National Science Foundation grant EAR-1057119 and by a UMN Grant-in-Aid to Dr. Jake Bailey. I would like to thank Rachel for her moral and emotional support, grammatical advice, immense patience, and advice that I couldn't have done without.

Abstract

As the study of microbes and their impact on the environment grows, so too does the desire to understand the genetic basis of the physiologies that make possible interactions between microbial cells and their environment. Since it is now much more cost-effective to sequence bacterial genomes, environmental metagenomic assembly is a very attractive option for obtaining the genetic blueprints of bacterial physiologies. Bacteria of the genus *Thiomargarita* (Greek; *theio*-: sulfur; *margarites*: pearl), pose a particularly interesting quandary. The genus includes the world's largest bacteria, but as uncultured organisms, their physiologies and basis for their gigantism are not well understood. In order to investigate the genetic basis for these modes, a single cell MDA amplification approach was used on *T. nelsonii* cells collected at the Hydrate Ridge methane seep off of the coast of Oregon. These particular cells were derived from a gastropod-attached epibiont community. Next-generation sequencing produced a metagenomic product representing both *T. nelsonii* and attached bacteria (epibionts). These reads were assembled into contigs, binned using the tetranucleotide frequency of the resultant contigs, and finalized using a more stringent secondary assembly. The resulting draft genome shows evidence in *Thiomargarita nelsonii* for a complete denitrification pathway not previously known in large, vacuolated, sulfur-oxidizing bacteria. Additionally, the genes necessary for polyphosphate metabolism were observed. Polyphosphate metabolism is thought to play a role in the formation of phosphatic minerals that serve as important reservoirs in the marine phosphorous cycle.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AOM	Anaerobic Oxidation of Methane
ATP	Adenosine Tri-Phosphate
Contig	Contiguous section of overlapping DNA that forms a consensus sequence
DMSO	Dimethyl Sulfoxide
DNA	DeoxyriboNucleic Acid
DNRA	Dissimilatory Nitrate Reduction to Ammonia
ESOM	Emergent Self-Organizing Map
GTE	GlycosylTransferase-Encoding
IMG-ER	Integrated Microbial Genomes Expert Review
ITS	Internal Transcribed Spacer
MDA	Multiple Displacement Amplification
MEGA	Molecular Evolutionary Genetics Analysis
MEGAN	Metagenome Analyzer
MSI	Minnesota Supercomputing Institute
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NCBI	National Center for Biotechnology Information
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
rDSR	Reverse Dissimilatory Sulfate Reductase
ROV	Remotely Operated Vehicle
rRNA	Ribosomal RiboNucleic Acid
SNP	Single-Nucleotide Polymorphism
(T)BLAST(N/P)	(Translated) Basic Local Alignment Search Tool (for Nucleotides/Proteins)
TCA	Tricarboxylic Acid
WGA	Whole Genome Amplification

1. Introduction

Deep-sea cold seeps exhibit both weak and ebullient flow of hydrocarbons such as methane that can vary (6-10^5 mmol m^{-2} day^{-1}) over spatial scales of only a few meters (Torres *et al*., 2002). The elevated flux of hydrocarbons from the subsurface in these settings drives the production of sulfide via the anaerobic oxidation of methane (AOM) and sulfate reduction (Boetius *et al*., 2000). Microbial mats at these sites, which can be spatially extensive, are visually dominated by large, sulfur-oxidizing bacterial ecotypes of the closely related genera *Beggiatoa* and *Thiomargarita* that fall within the family *Beggiatoaceae*. These bacteria are chemolithotrophs that obtain energy for metabolism from the oxidation of reduced sulfur species. *Thiomargarita* are thought to primarily use the oxidation of electron donors available in the sediment to fuel carbon fixation. However, the terminal electron acceptors used in these reactions can vary depending on their location relative to the oxic/anoxic boundary. Under these highly reducing conditions, oxygen concentrations decline rapidly upon penetration into the sediments. Where sulfide fluxes upward into the water column, the oxic/anoxic boundary changes temporally and spatially, forcing bacterial species to adopt alternate electron acceptors to survive the dynamic conditions. Nitrate can be used as a terminal electron acceptor in large, vacuolated sulfur bacteria in times of anoxia; *Thiomargarita* allocates up to 90% of its volume for intracellular nitrate storage (Schulz and Jørgensen, 2001). *Thiomargarita* has also been shown to accumulate intracellular elemental sulfur inclusions that serve as intermediates in the oxidation of hydrogen sulfide to sulfate, to provide the cell with electron donor when access to sulfide is limited. The lack of motility observed in *Thiomargarita* additionally illustrates the need for metabolic flexibility in the presence of temporally and spatially variable geochemical gradients. These nitrate and sulfur storage capacities allow them to bridge the suboxic zone, where neither sulfide nor oxygen exist, providing *Thiomargarita* an advantage over sulfide-oxidizing bacteria that are incapable of spanning this gap (Schulz *et al*., 1999). *Beggiatoa* and *Thioploca* have been shown to possess

mechanisms for motility across the suboxic zone, with *Thioploca* cells spanning the entire region, and *Beggiatoa* following the oxygen gradient both above and below the sediment surface (Schulz & Jørgensen, 2001).

Prior research has demonstrated that *Thiomargarita* is capable of accumulating phosphate intracellularly as long polyphosphate (poly-p) polymers. The hydrolysis of this polyphosphate, and concomitant release of phosphate into pore water has been linked to the formation of large phosphorite deposits in the seafloor and the subsurface (Schulz & Schulz, 2005; Bailey *et al*., 2007; Goldhammer *et al*., 2010). These phosphatic minerals precipitated by this release of phosphate serve as a form of sequestration, or long-term phosphorous sink, indicating that the role of the large sulfur bacteria in the phosphorous cycle is worth investigation.

Thiomargarita-like bacteria have been discovered in numerous locations on the ocean floor, including the Gulf of Mexico (Kalanetra *et al*., 2005; Bailey *et al*., 2007), mud volcanoes in the Barents Sea (Girnth *et al*., 2011), the Mediterranean Sea (Grünke *et al*., 2011), the Costa Rican Margin (Schulz *et al*., 1999), and methane seeps off the coast of Oregon. The "sulfur pearl" *Thiomargarita* (Greek; *theio*-: sulfur; *margarites*: pearl) species initially discovered at seeps along the Costa Rican Margin and later observed at Hydrate Ridge, *T. nelsonii*, were found commonly attached to substrates, in particular on shells of *Provanna* snails. Additionally, these attached cells appeared to undergo an apparent dimorphism (elongate vs. budding) in their life cycle, wherein *Thiomargarita* elongated almost a millimeter in length, and budded into spherical, or coccoidal, daughter cells on the free end (Bailey *et al*., 2011). Several of the samples collected showed a distinct epibiont community resembling morphotypes similar to *Ca. Thiomargarita nelsonii* as well as *Maribeggiatoa* sp. and *Leucothrix* sp. Due to prior failed attempts to culture *Thiomargarita* (Flood & Bailey, unpublished), a bioinformatics-based *de novo* sequencing and assembly approach was chosen to reveal the genetic information coding for metabolic enzymes, cytoskeletal proteins, and the extent of lateral gene transfer in *Thiomargarita*. Certain

Thiomargarita nelsonii morphotypes exhibit a dimorphic reproductive behavior not previously described in other large sulfur bacteria (Bailey *et al*., 2011). Selection was therefore made of both bud and stalk morphotypes for amplification and sequencing. Despite attempts at single-cell isolation via physical cell straining, PCR screening revealed that all 4 collected samples contained other bacterial cells in addition to the targeted *T. nelsonii*. As the amplification of the *Thiomargarita* 16S and ITS regions produced cleaner Sanger sequencing chromatograms in the coccoidal buds, the genome amplification, sequencing and assembly was only carried out on the coccoid buds. It is thought that the elongate cells not sequenced here experience a larger breadth of geochemical conditions as compared to their stationary mat-forming counterparts, as attachment to their mobile host provides transportation.

None of the vacuolate sulfide-oxidizing bacteria have been isolated in pure culture. Thus, what is known about their unique morphology, their nitrate storage abilities, and their ecological relationships is primarily based on extrapolation of isolated bacteria with seemingly similar characteristics. Previous physiological and genetic studies have been carried out on *Beggiatoa* sp. (Nelson & Jannasch, 1983; Ahmad *et al*., 2006; Mußman *et al*., 2007; MacGregor *et al*., 2013), a closely related large sulfur-oxidizing bacteria. These studies showed the presence of an adaptable suite of metabolic-enzyme-coding genes that allow *Beggiatoa* to persist in habitats that would be too inhospitable for other bacterial species without said mechanisms. The primary line of investigation into the similar ecophysiological role of *Thiomargarita* in its locally dynamic environment was to sequence and investigate the genes involved in metabolism of the chemical species present. Investigations of *Thiomargarita* sp. have been centered around physiology studies, and no studies of the genome have yet been published.

It is now standard to study genomic fragments of uncultured microbes obtained by shotgun sequencing of bulk DNA extracted from mixed communities (Tyson *et al*., 2004; Dick *et al*., 2009). However, the assembly of whole genome amplification (WGA) from metagenomic

datasets is a technique that is still in its infancy, and has been shown to be problematic with respect to genome coverage. Multiple displacement amplification (MDA), a type of whole genome amplification (WGA), can amplify bacterial DNA (up to a billion-fold) from single cells or mixed communities, can generate higher-quality sequences from uncultured environmental microorganisms than previous WGA techniques (Lasken & Stockwell, 2005; Raghunathan *et al*., 2005; Kvist *et al*., 2007; Podar *et al*., 2007). However, this technique does have disadvantages, as it has been shown to create segments of DNA that have no corresponding representatives in the environment, essentially rearranging sections of genetic code to produce what are known as chimeras, or false combinations of two or more pieces of DNA. The amplification of the DNA that is separate from the target DNA, or overrepresentation of bacterial DNA in species that are relatively rare in the environment has also been shown to confound analysis of microbial genomes (Binga *et al*., 2008). Previous methods to extract microbial genetic information from mixed communities were unable to amplify whole genomes, as the stochastic nature of amplifications would commonly amplify only the most abundant species. Despite background amplification and chimera formation (Lasken & Stockwell, 2005), MDA has higher efficacy in amplifying complex DNA than past WGA techniques (Gonzalez et al., 2005; Spits *et al*., 2006). One of the ways to combat these errors in amplification is to use high-throughput Illumina sequencing. This type of sequencing generates billions of short (100 base pair) sections of genetic information, wherein overlapping segments of DNA can be assembled into long sections of consensus. In addition, paired-end reads contain positional information, allowing for highly precise alignment of reads over many times the length of the reads themselves. This positional information, paired with the consensus sequences generated from assembly, obviates some of the problems associated with chimera formation from amplification.

The combination of the whole-community representation of MDA-amplified DNA with the coverage and positional advantages of Illumina paired-end sequencing show great potential to rapidly analyze the genomes of unculturable microbes. Using MDA, the genomic DNA of two

cells of uncultured *Thiomargarita* from a bacterial colony were separately amplified in this study. Both of these cells were sequenced on Illumina HiSeq lanes, and annotated using the Integrated Microbial Genomes (IMG) Pipeline. Here I summarize the approach of constructing a draft genome of the world's largest bacteria from an environmental sample, highlight the metabolic pathways inferred from genes in this genome, and discuss the potential role of those genes in *Thiomargarita nelsonii*.

2. Methods

2.1. Sample Collection

The Remotely Operated Vehicle (ROV) *Jason* was used to collect *Provanna* sp. snail samples during a research expedition on board the R/V *Atlantis* (AT18_10) to Hydrate Ridge North (44º 40.02687' N 125º 5.99969' W), Cascadian margin, off the coast of Oregon, USA on September 3, 2011. The shells of live *Provanna* sp. snails collected from the methane seep sediments hosted an attached microbial (i.e., epibiont) community that appeared to be rich in sulfide-oxidizing bacteria (Fig. 1), including morphotypes that resembled *Candidatus Thiomargarita nelsonii,* as well as *Ca. Maribeggiatoa* sp. and *Luecothrix* sp. Some snails were preserved shipboard individually in a 1:1 mixture of sterile ethanol and Instant Ocean (Instant Ocean, Blacksburg, VA) and then stored at -20°C for future molecular analyses. Live snails, as well as snails fixed in 4% paraformaldehyde diluted in Instant Ocean, were also maintained and later observed in the laboratory. An Olympus SZX16 stereo-microscope and an Olympus BX61 compound microscope equipped with an Olympus XM10 color camera were used to obtain high-resolution images of the snails and the attached cells.

2.2. Genomic DNA Preparation and Sequencing

Stereomicroscopy revealed coccoidal buds either partially or fully detached from the elongated *Thiomargarita nelsonii* morphotypes in the ethanol-seawater mixture surrounding snail samples. These individual *Thiomargarita* cells were then pipetted into a sterile 40 μm pore-diameter cell strainer (BD Biosciences, San Diego, CA) within a well of a sterile 6-well plate. Each well contained DNA-free 3.5% NaCl and 50% ethanol. Using aseptic techniques, the cell strainer was then lightly gyrated to disassociate small unattached and loosely attached bacteria from the *Thiomargarita* cells. Then the cell strainer was transferred to a clean well and the process was repeated five times. The cell strainer was then placed over a 50 mL conical and ~10 mL of the saltwater-ethanol mixture was gently poured over the cell strainer. The cell strainer was then placed into a DNA-free phosphate buffer solution. Individual *Thiomargarita* cells were then transferred by pipette to 5 μL of the PBS solution and frozen separately at -20°C.

Two coccoidal buds and two elongated cells were selected for genomic DNA amplification. DNA extraction and amplification was performed using the Qiagen REPLI-g Midi kit (Qiagen, Hilden, Germany) for whole genome amplification (WGA), following the manufacturer's protocol. Cell lysis was achieved by heating to 65ºC and using the proprietary Qiagen lysis buffer D2. One coccoidal and one elongated cell were also subjected to an additional step of 10 minutes of incubation on ice following the cell lysis step, to determine if additional cooling step increased genetic yield. The DNA was eluted in Tris-EDTA and an 1:100 dilution aliquot was generated for DNA quantification via a Thermo Scientific Nanodrop 2000C Spectrophotometer and for polymerase chain reaction (PCR) screening. The undiluted stocks developed a white fluffy precipitate (presumably proteins and other biomass). Thus the stocks were centrifuged at 10,000 rpm and the more pristine supernatant was removed and used as the primary DNA stock for genomic sequencing.

PCR utilizing the primer set 233F-ITSReub was utilized to confirm positive amplification of *Thiomargarita nelsonii.* The samples were screened for contamination using the universal

primers for the 16S gene for bacteria (27F-1492R) and archaea (Arc8F - 9rts) (Table 1). Agarose gel analyses and Sanger sequencing revealed that all four samples contained bacterial DNA but none contained archaeal DNA. The elongated cells showed less amplification of the *Thiomargarita* 16S and internal transcribed spacer (ITS) region and appeared more contaminated with environmental bacteria than the coccoid buds and thus were excluded from further analyses. Stock DNA was submitted to University of Minnesota Genomics Center for DNA quality screening and genomic sequencing. Sample B6 was sequenced on a full lane of a Illumina HiSeq 2000 (Illumina, San Diego, USA) with 100 bp paired-end reads with ~300 base inserts while sample B10 was sequenced on a half a lane. Using the Lander/Waterman equation (Lander & Waterman, 1988) for computing coverage, (L=100 bps, N=1.975x10^8 reads, G=7.4x10^6 bps) if the sample was uncontaminated, we would expect ~2600x coverage, with the genome length equaling the length of the *Beggiatoa* sp. PS genome (Mußmann *et al*., 2007). Given the fact that the extract contained DNA from multiple organisms, this estimate served as an extreme upper bound on sample coverage. The presence of 20+ distinct 16S rRNA gene sequences in the raw dataset indicates that a more accurate upper coverage bound is ~130x, based on an average microbial genome size equivalent to *Beggiatoa* sp. Due to the high number of genome copies, or polyploidy, observed in large bacterial cells (Mendell *et al*., 2008; Viswanathan, 2012) an overrepresentation of genetic material from *Thiomargarita* is likely responsible for higher representation in the environmental sample. Dichosa *et al*. (2012) were able to show that inducing artificial polyploidy in *Bacillus subtilis* cells markedly increased sequencing coverage, illustrating that genome copies can improve assembly quality.

2.3. Bioinformatic Analysis

Bioinformatic analysis was performed using supercomputing resources from the Minnesota Supercomputing Institute (MSI), the Michigan Geomicrobiology Lab, and Symbiosys, the Bailey Geobiology Lab server. Initial quality analysis was performed using Galaxy (Giardine *et al*., 2005; Blankenberg *et al*., 2010; Goecks *et al*., 2010) through MSI, using the FASTQC

module (Andrews, 2010). FASTQC-MCF was used to filter out low-quality sequences (>25 quality score) and remove adapters left over from Illumina sequencing (Aronesty, 2011). PRINSEQ (PRINSEQ lite 0.19.3) was also used to trim low quality ends (>5 base pairs) from otherwise high-quality sequences (Schmieder & Edwards, 2011). A custom-made Perl script was used to remove duplicate sequences from the trimmed and filtered sequences (Jones, unpublished). These analyses, unless otherwise noted, were performed on the MSI's "Labs" server, as well as on their high performance computing cluster, Itasca (Table 2).

A separate protocol of quality analysis was carried out on the samples processed by the Michigan Geomicrobiology Lab. A Perl script designed to remove any raw reads with 100% identity over 100% length was first used to remove exact duplicates, reduce file size and reduce assembly time. Trimming was performed using Sickle with default parameters, on the forward and reverse reads separately. The trimmed and dereplicated reads were then interleaved using a custom Perl script, and converted to a .fasta file for assembly. Assemblies were made using the String Graph Assembler (SGA), ABySS-pe, IDBA-UD, and MetaVelvet (Table 3) (Simpson & Durbin, 2010; Simpson *et al*., 2009; Peng *et al*., 2012; Namiki *et al*., 2012; respectively) with k-mer lengths greater than 65% of raw read length (>65 bps). K-mer lengths were chosen to maximize computational and assembly efficiency. A broad variety of assemblers were tested in order to determine the most effective assembler for assembling genomic data from this mixed-community sample. Outputs of assembled contigs were visualized using the Metagenome Analyzer (MEGAN) to determine phylogenetic affiliation (Huson *et al*., 2011), and additionally visualized with Hawkeye and the ABySS explorer to determine coverage (Schatz *et al*., 2013).

MEGAN and the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool for Nucleotides (BLASTN) were used at a minimum cutoff of 1e-10, to determine which assembler produced the longest contigs that hit to the Thiotrichales order (to which *Thiomargarita* belongs). As shown in Table 3, IDBA-UD appeared to produce the best

assembly by many metrics, including number of contigs generated, number of contigs over 1000 base pairs, mean contig size, and N50 length. However, analysis of the contigs belonging to *Thiotrichales* showed that a large amount of misassemblies were produced with this assembler, a problem not observed in MetaVelvet, a more conservative assembler designed to assemble metagenomic data. Once MetaVelvet was found to produce the longest *Thiotrichales* contigs, future assemblies were carried out with MetaVelvet exclusively. Three assemblies were done at k-mer lengths of 53, 71, and 89. The assemblies were combined using Minimus2 (Sommer *et al*., 2007), an assembler in the AMOS package, with minimum overlap settings of 100 base pairs, and minimum identity percentage of 99%. These settings were chosen in order to minimize chimeric assemblies. A tetramer-frequency based Emergent Self-Organizing Map (tetra-ESOM) was constructed on the Geo-omics server in the Michigan Geomicrobiology Lab following the protocol outlined in Ultsch & Mörchen (2005) and Dick *et al*. (2009). The protocol that was followed can be found at (https://github.com/tetramerFreqs/Binning). After the initial creation of the necessary files for the SOM to run, the default parameters were selected for the ESOM run. The network was trained with a K-Batch algorithm, 220 rows and 220 columns (~48400 neurons), and a starting radius value of 50. After binning by the tetra-ESOM method, BLASTN was used to query the 16S and 23S rRNA sequences characteristic for *Ca. Thiomargarita nelsonii* and closely related *Thiomargarita sp.* against the emergent bins. CAP3 (Huang & Madan, 1999) was used to perform a final assembly with a minimum overlap of 150 base pairs on the sequences in the bin that was identified via BLAST as having originated from *Thiomargarita nelsonii*.

2.4. Annotation and Function Mapping

Annotation was carried out initially by the Integrated Microbial Genomes Expert Review (IMG-ER) automated gene-calling pipeline (Markowitz *et al*. 2012). Manual curation of the relevant metabolic and structural genes was carried out using the IMG BLAST function and neighborhood operon analysis. In the absence of clearly annotated function from the IMG Pipeline, individual genes were sought out using a best BLAST match from closely related

organisms (*Thiomicrospira*, *Beggiatoa* sp. PS, *Beggiatoa* sp. Orange Guaymas). After an initial protein-protein BLAST (BLASTP) of the gene in question, any genetic elements selected as potentially missed by the IMG Pipeline were queried using translated nucleotide BLAST (T-BLASTN) against the NCBI BLASTN database. A minimum e-value cutoff of 1e-100 and minimum percent identity of 65% were used to determine annotation at the protein level.

3. Results

3.1. Sequence Binning

In order to identify and isolate the genetic material of *Thiomargarita* from sequences originating in other bacteria from the mixed community, all contigs of our MetaVelvet-assembled dataset were analyzed in a binning approach based on their intrinsic DNA codon usage bias. The relative abundance of tetramers was used to statistically train an Emergent Self-Organizing Map (ESOM) to cluster closely related sequences, as has been shown to be effective in separating even closely related bacterial species from a metagenome (Dick *et al*. 2009). Fig. 2A shows the resulting ESOM with sequences ranging from 2.5-5 kbps, with the *Ca. Thiomargarita nelsonii* 23S rRNA sequence used to identify the bin of interest (labeled here as Bin 1). This range was chosen to extract the highest degree of separation between bins, as shown by the brown-colored gaps between extracted bins. Fig. 2B shows the ESOM that was created with a range of 4-8 kbps, with none of the bins strongly affiliated to the genome in question. The resulting bins warrant further consideration as they are likely ecologically-relevant epibionts, with the labeled Bin 2 (Fig. 2C) identified as a Deltaproteobacterium, some of which are known to reduce sulfur.

3.2. Phylogeny and Morphology

In both spherical (coccoidal) buds that were sequenced and analyzed, partial 16S rRNA gene sequences were found that were identified as *Ca. Thiomargarita nelsonii*. Comparative analysis of the 16S rRNA sequences and ITS regions revealed a 100% identity match between the

buds along the entirety of the present 16S and ITS sequences, indicating that the separate buds are of the same strain of *Ca. Thiomargarita nelsonii*. This is perhaps unsurprising given that the cells originated from the same snail shell.

Fig. 3 shows a maximum likelihood tree of the 16S rRNA gene of the *Beggiatoaceae, Leucotrichaceae,* and *Achromatiaceae* in relation to the morphology of the more recently sequenced organisms in bold (Salman *et al*. 2013). The sequenced buds matched with 100% identity to the 16S fragments of *Ca. Thiomargarita nelsonii* HYR001 and COS (008, 015, 013, 010, 012), with 100% match of the ITS region fragment to the HYR001 phylotype. A comparative geographic morphotype analysis as shown in Fig. 4 illustrates that the dimorphic bud and stalk morphotypes present in *Thiomargarita* from Hydrate Ridge correlate to the 16S rRNA similarity observed in this study (Bailey *et al*. 2011).

3.3. Gene Identification

Beggiatoa sp. *Orange Guaymas* and *Beggiatoa* sp. PS are the closest relatives of *T. nelsonii* for which whole genome sequences are available on the Integrated Microbial Genomes (IMG) database, based on 16S rRNA similarity. A comparison of the conserved open reading frames (ORFs) (Table 4) present in the genome against the IMG database supports this affiliation, as best matches showed highest degrees of similarity to members of the *Beggiatoaceae*.

Many genes identified as belonging to *Thiomargarita* showed closest phylogenetic similarity to cyanobacterial genes from the filamentous *Nostoc* sp. and *Cylindrospermum stagnale*, and the gliding cyanobacterium, *Anabaena variabilis*. Most of these ORFs encode genes relevant to cell membrane biogenesis, transport of secondary metabolites, and conserved hypothetical proteins (Table 5). Similar phylogenetic patters for hundreds of genes in *Beggiatoa* were interpreted by Mußmann *et al*. (2007) and Flood *et al*. (2014) to result from horizontal gene transfer. The phylogenetic clustering of genes found within the *Thiomargarita* genome to genes

present in cyanobacteria (instead of proteobacteria) suggests that horizontal gene transfer is the likely mode of acquisition, in contrast to descent with modification. Extensive gene exchange was shown between filamentous cyanobacteria and *Beggiatoa* sp. as the putatively transferred genes co-localized with the nitrate reductase subunit genes native to *Beggiatoa*. Additionally, contigs with cyanobacteria-affiliated genes did not group in their cluster analysis, furthering the hypothesis of horizontal transfer. In our tetra-ESOM binning method, those contigs containing these putatively transferred genes also did not group together, which indicates an already *Thiomargarita*-adapted codon usage pattern. In Fig. 5, a phylogenetic tree was constructed to show the relationship between a putatively horizontally-transferred glycosyltransferase (Tmarg_03686) and the top 20 closest protein BLAST (BLASTP) results. An NCBI BLASTP search was performed using the default scoring matrix (BLOSUM62) and the default e-value of 10. The top 20 BLAST results were aligned using the Molecular Evolutionary Genetics Analysis (MEGA 5.22) software package (Tamura *et al*., 2011). Sequences were aligned using the MUSCLE (MUltiple Sequence Comparison by Log-Expectation) (Edgar, 2004) alignment module built in to MEGA 5.22. Default alignment settings within the MUSCLE alignment module were used. The phylogenetic tree was constructed using a maximum-likelihood tree-building method with a 1000 replicate bootstrap analysis. Fig. 5 shows that most representatives belong to the Cyanobacteria or Planctomycetes, with *Thiomargarita nelsonii* falling into the middle of the cyanobacterial grouping.

Additionally, genes were identified in the binned *T. nelsonii* that clustered most closely to genes from Alpha-, Beta-, and Deltaproteobacteria, represented most commonly by *Rhodobacteraceae*, *Commonadaceae*, and *Desulfobacteraceae*, respectively. Most of the contigs were identified as conserved hypothetical proteins, with glycosyltransferases and other cell membrane biogenesis proteins present in all groups, exhibiting particularly strong (>60% percent identity) matches to the Betaproteobacteria (Table 6). Putative transfers from the Betaproteobacteria also include transposases (Tmarg_11326, Tmarg_11829), or enzymes that

serve to shuttle mobile genetic elements to another part of the genome by either a cut-and-paste insertion mechanism or a replicative transposition.

A heat map of glycosyltransferase-encoding genes was constructed from the genome of *Thiomargarita nelsonii*, the closely-related *Beggiatoa* sp. Orange Guaymas, *Beggiatoa* sp. PS, the chemolithotrophic, sulfide-oxidizing bacteria *Thiomicrospira crunogena* (1-2 μm cell diameter), and the Gammaproteobacteria *Vibrio harveyi* (Table 7). The large, vacuolated *Beggiatoaceae* and *Thiomargarita* species reveal a large amount (up to 37 in *Thiomargarita*, 19 in *Beggiatoa* sp. Orange Guaymas) of copies of the glycosyltransferase-encoding (GTE) genes. The *Thiomargarita* genome contains nearly double of GTE genes in previous studies of large sulfur-oxidizing bacteria (Mußmann *et al*., 2007; MacGregor *et al*., 2013).

3.4. Ribosomal Proteins and Single-Copy Genes

To determine the completeness of our assembled genome, the presence and number of duplicate genes that usually only occur once per prokaryotic genome were analyzed. Forty-one ribosomal proteins in the *T. nelsonii* dataset were identified that exclusively affiliated with Gammaproteobacteria, 7 of which identify at greater than 90% sequence similarity to the aforementioned class (Table 8). Raes *et al*. (2007) proposed a novel approach for the prediction of the number of genome equivalents in metagenomic samples, that is based on the occurrence of 35 widely conserved, single-copy marker genes present in most prokaryotic genomes. All 35 of these genes were identified in the *T. nelsonii* dataset, some genes more than once (Table 9). These duplicates could be a result of an artifact of the tetra-ESOM binning approach and the stringent minimus2 meta-assembly parameters (OVERLAP=100, MINID=99); an indicator of single-nucleotide polymorphisms (SNPs) among the many genome copies that *Thiomargarita* is known to possess; or a product of amplification, from the polyploidy shown in *Thiomargarita* or from MDA. The presence of all 35 single-copy marker genes as determined in Raes *et al*. (2007) indicates the relative completeness of the *T. nelsonii* genome. These single-copy genes share

extremely high degrees of similarity between duplicates, not only in operon makeup and function, but high levels of nucleotide-level similarity (>99% identity).

3.5. Nitrate Respiration

Analysis of the *T. nelsonii* dataset reveals two complete denitrification pathways. One involves a nitrite reductase (*nirBD*, Tmarg_08782, Tmarg_08783) that catalyzes the final step in the reduction of nitrate to ammonium (DNRA), while the other involves a nitrous oxide reductase (*nosZ*, Tmarg_08173-4) that catalyzes the final step in denitrification (Fig. 6A). The nitrate reduction pathway encompasses both membrane-bound (*narGHIJ*, Tmarg_01615-01618, Tmarg_11764-11767) and periplasmic (*napAGH*, Tmarg_12325-12327, Tmarg_06472-06478) nitrate reductases. The role of NapAGH in the reduction of nitrogen is unclear, but it has been proposed that it may allow *T. nelsonii* and *Beggiatoa* sp. to support nitrate respiration at low nitrate concentrations (Wang *et al*., 1999) or enable respiration of nitrate under anaerobic conditions (Bell *et al*. 1990). A nitrite reductase (*nirS*, Tmarg_03321, Tmarg_07219, Tmarg_12671) and two nitric oxide reductases (*norBC*, Tmarg_08828, Tmarg_03322) reduce nitrite to nitric oxide and nitric oxide to nitrous oxide, respectively.

3.6. Sulfur Oxidation

The genomes of the coccoidal buds sequenced encode proteins of the reverse dissimilatory sulfate reductase (rDsr) pathway (Hipp *et al*., 1997; Pott & Dahl, 1998) (Fig. 6B). Gene fragments were identified that encode the cytoplasmic DsrABC (*dsrABC*; Bud 2) and the membrane proteins DsrJKMOP (*dsrKMJ*; Tmarg_08388, Tmarg_08387, Tmarg_08389; *dsrOP*; Bud 2) that serve to funnel electrons to DsrABC. Similar to the betaproteobacterium *Thiobacillus dentrificans* (Beller *et al*., 2006) and *Beggiatoa* sp. (Mußmann *et al*., 2007), at least 3-5 homologs of the DsrC-like subunit are present in the *T. nelsonii* genome. Following the formation of sulfite by the DsrABC complex, it is oxidized and phosphorylated by an adenosin-phosphosulfate (APS)

reductase to APS (*aprAB*; Tmarg_02567-02568). APS is then dephosphorylated with an ATP sulfurylase (Tmarg_05810, Tmarg_00628) to produce sulfate and an ATP (Hagen & Nelson, 1997). In *Beggiatoa* sp. and *T. nelsonii*, the AprAB is linked to heterodisulfide reductases (*hdrABC*; Tmarg_02860-02862) that transport electrons to AprAB, as has been shown in sulfate-reducing bacteria (Mußmann *et al*., 2005, Haveman *et al*., 2004). Analysis revealed the presence of a gene encoding sulfite oxioreductase (*sorA*; Tmarg_09133), which oxidizes sulfite directly to sulfate.

The oxidation of thiosulfate is catalyzed in *T. nelsonii* by the SoxABXYZ (*soxABX*, Bud 2; *soxYZ*, Tmarg_02753-4, Tmarg_06834-5) subunits of the SOX pathway (Friedrich *et al*., 2001). As has been shown in *Beggiatoa* sp. and *Allochromatium vinosum*, some organisms that encode the rDsr pathway form sulfur globules from the oxidation of reduced sulfur compounds (Mußmann *et al*., 2007, Hensen *et al*., 2006), likely related to the lack of SoxCD genes. The products of the SoxCD are responsible for the formation and release of an additional sulfate from the SOX system, yet is not required for sulfite oxidation (Friedrich *et al*., 2001) *Thiomargarita nelsonii* lacks these SoxCD subunits in both currently sequenced buds, and is known for its production of elemental sulfur globules. It is therefore not expected that the SoxCD genes would be present in any of the unsequenced portions of the genome.

The adaptability of the respiratory pathways of *T. nelsonii* is demonstrated by the presence of dimethyl sulfoxide (DMSO) reductase genes (*dmsABC*, Tmarg_02306-8) indicating that DMSO can be respired in addition to nitrate or oxygen. DMSO is formed by eukaryotic plankton burial (Besiktepe *et al*., 2004) and photochemical oxidation of dimethyl sulfide (Brimblecombe & Shooter, 1986). Therefore, in anoxic, nitrate-poor waters, *T. nelsonii* could access this electron acceptor at the sediment-water interface. Furthermore, the presence of thiosulfate reductase (*phsAC*, Tmarg_01089, Tmarg_07692) genes in the *Thiomargarita* genome

indicates a metabolism that could be involved in the reduction of elemental sulfur and tetrathionate, as has been proposed in Hinsley & Berks (2002) and Mußmann *et al*. (2007).

3.7. Phosphate Accumulation

Under nutrient stress, many bacterial species will accumulate intracellular phosphate stored as polyphosphate, long chains of phosphate joined by shared oxygen atoms. Large, vacuolated sulfur bacteria have been shown to exhibit this phosphate storage and contain large polyphosphate granules (Schulz & Schulz, 2005). Both *Thiomargarita* datasets encode for phytases (Tmarg_01186), enzymes critical to the uptake of inorganic phosphates from the sediments known as phytates. As has been shown in related species (Mußmann *et al*. 2007), *T. nelsonii* takes up orthophosphate and polyphosphate from the environment by way of phosphate permeases, symporters, and phosphate-selective porins (Tmarg_02751, Tmarg_06511). Poly-p granules are synthesized via poly-p kinase (*ppk*, Tmarg_11806, Tmarg_02749, Tmarg_08103) (Fig. 6C), from phosphate taken from ATP. In addition, genes were found that encode polyphosphate pyrophosphohydrolases/synthetases (Tmarg_08051, Tmarg_11732, Tmarg_04433, Tmarg_07574) that serve to catalyze the release of phosphate groups from polyphosphate to aid in purine metabolism in other bacteria. It is likely that the hydrolysis and release of this phosphate occurs only in times of nutrient stress, when *Thiomargarita* is unable to uptake electron acceptor from the surrounding pore water or sediments. The oxic-anoxic boundary may serve as a dividing line between accumulating poly-p in aerobic sediments and degrading poly-p under anaerobic conditions where acetate is present (Karl, 2014).

3.8. Oxygen Respiration

The presence of both the high-affinity cytochrome c bb_3 (Tmarg_00823-00825, Tmarg_12758-12760, Tmarg_08520-08522) and the low-affinity cytochrome c aa_3 (quinol oxidase, Bud 2) (Fig. 6D) in the annotated genome demonstrates the flexibility of *T. nelsonii* to respond to both microoxic and high-oxygen conditions, respectively. The differential expression

of cytochrome oxidases under varying oxygen concentrations has been demonstrated for *Beggiatoa leptomitiformis* and *Beggiatoa* sp. (Muntian *et al*., 2005; Mußmann *et al*., 2007). *Beggiatoa* and their close relatives have been shown to exhibit negative chemotactic responses to high oxygen concentrations, and preferentially oxidize sulfur compounds in microoxic incubations (Møller *et al*., 1985), which could account for the observed trend of a higher number of gene copies that encode the cytochrome c bb_3 in the *T. nelsonii* genome.

4. Discussion

4.1. Sequencing, Analysis, and Binning

The assembly of raw, mixed community genetic data has been shown to be a challenging problem in microbiology and bioinformatics. The degree of similarity observed in genes from an environmental sample, some belonging to organisms that are very closely related, can create misassemblies that further obfuscate downstream analysis of assembled genetic product. A conservative protocol is necessary for the particular type of assembly and binning that must occur to bin out genomes from a metagenome. An assembler created with the intention of cautiously assembling the genetic material from only one phylotype is therefore imperative to later binning success. Many types of assemblers were tested, in order to determine maximum efficiency and effectiveness. A few types of assemblers noted for the low amount of computational resources required were utilized (String Graph Assembler, AbYSS-PE, SPAdes, IDBA), but as these algorithms were developed with single-cell sequencing in mind, the variable coverage found in our dataset caused a product with many short contigs (~150 bps). A metric consisting of a combination of the average length of assembled contigs, the number of contigs generated longer than 1000 base pairs, and the number of contigs generated was used to evaluate assembly efficacy (Table 3). Assemblers specifically designed to combat the uneven coverage (IDBA-UD, MetaVelvet) were tested next, with MetaVelvet (Namiki *et al*., 2012) producing similarly sized contigs, while creating the few misassemblies. IDBA-UD has a limited scaffolding capability, and

has been shown to exhibit lower coverage at every taxonomic level than MetaVelvet, and lower capability to extract coverage for genomes with a low representation in a metagenomic sample than MetaVelvet (Namiki *et al*., 2012). MetaVelvet was therefore chosen as the assembler of choice after comparison with other protocols (Table 3).

Binning of the *Thiomargarita* draft genome from our assembled dataset was first carried out using principal component analysis (PCA) on a matrix of tetranucleotide frequencies inherent to the contigs generated prior. The relative frequency of these tetramers across contigs can be used as a fingerprinting of genetic material, a way to cluster genes with similar codon-usage bias together. After PCA was unable to show meaningful separation among supplied contigs, we attempted the Emergent Self-Organizing Map protocol outlined in Dick *et al*. (2009) to make use of a machine-learning approach to cluster our contigs based off of their signature tetranucleotide frequency. This technique uses a neural network algorithm to represent multidimensional data on a two dimensional map, with map 'elevation' corresponding to distance in tetranucleotide frequency between contigs. This approach was largely successful, an ESOM generated with contigs between 2.5 and 5 kbps possessed a phylogenetic marker gene belonging to *Thiomargarita nelsonii*, and showed separation from the background genetic material. Due to the human-aided boundary markings made on our ESOM, additional analysis was necessary to confidently assign microbial taxa to the produced bins. Contigs at the edges of the bin were shown to belong to bacteria other than Thiomargarita that were present in the original metagenome, and a manual curation of the annotated *Thiomargarita* bin was carried out to ensure the highest quality of the draft genome. The ESOM protocol used should not stand alone as the sole binning method, as alternate methods should always be used to corroborate any assertions regarding assigned microbial taxa due to the subjective nature of delineating the bin boundaries. It is possible that genes from organisms with similar codon usage biases are present in the draft genome, yet those genes essential to metabolism have been identified as belonging to *Thiomargarita* via best BLAST match analyses to closely related organisms. In addition, genes that have been recently horizontally transferred into the *Thiomargarita* genome and not yet

adopted its codon usage bias, or similar tetranucleotide frequency could also be missing from the draft genome. Investigations into additional binning strategies (i.e., coverage-based binning) will be carried out in the future, to provide a supplemental genetic clustering mechanism for analysis.

4.2. Site and Geochemistry

The microbial ecology of Hydrate Ridge, with its complex and variable geochemistry, invites further exploration. Hydrate Ridge, in the Cascadia margin accretionary complex off of the coast of Oregon, USA, is characterized by extensive deposits of methane hydrate. The ridge itself has a northern topographic high located at (44º 40.02687' N 125º 5.99969' W) at about 600m water depth, which is comprised of extensive carbonate deposits, and a southern maximum (44° 34.1' N 125° 9.0' W) at about 800m depth which is primarily covered in sediments. The entirety of the range has been described as a complex hydrogeologic system wherein fluid and methane fluxes vary by up to five orders of magnitude across the seafloor (Torres *et al*., 2002; Tryon *et al*., 2002). There are three active fluid flow regimes present at this site, discrete sites of methane ebullition driven by destabilization of gas hydrate deposits, large bacterial mats capping methane hydrate crusts, and sites colonized by vesicomyid clams (which harbor sulfide oxidizing bacteria in their gills) characterized by orders of magnitude slower gas release (Torres *et al*., 2002). The bacterial mats at the northern maximum cover areas as large as 15-20 m^2 and the underlying sediment has been shown to release methane at rates ranging from 30-100 mmol m^{-2} day^{-1} (Tryon *et al*., 2002; Boetius & Suess, 2004), compared with 10^6 mmol m^{-2} day^{-1} methane flux at discrete release points nearby.

Sahling *et al*. (2002) were able to show that the distribution of benthic communities observed at Hydrate Ridge are related chiefly to the flux of sulfide from the surface sediments. In turn, the sulfide fluxes are regulated largely by the supply of methane from the underlying sediments and sulfate from the water column. Methane and sulfate serve as electron donor and

electron acceptors, respectively, for anaerobic oxidation of methane (AOM) mediated by an active biological community of methanotrophic archaea and sulfate-reducing bacteria (Boetius *et al*., 2000; Nauhaus *et al*., 2002; Treude *et al*., 2003; Joye *et al*., 2004; Bowles & Joye, 2011; Green-Saxena *et al*., 2013). Production of sulfide at this interface coincides with vast mats of large, vacuolated, sulfide-oxidizing bacteria, both *Beggiatoa* sp. and *Thiomargarita* sp. The sulfur gradient underneath the bacterial mats measured by Boetius & Suess (2004) show sulfide increasing from 0 mM at the sediment-water interface to 10 mM at 5 cm depth, while sulfate decreases from 28 mM at the sediment surface to 3 mM at 5 cm depth before rising slightly. At the sites sampled with *Beggiatoa* sp., the sulfide is less concentrated than in sediments not covered with bacteria, despite significantly higher rates of sulfate reduction. This is likely due to the higher rate of sulfide oxidation correlated to bacterial metabolism (Knittel *et al*., 2003; Boetius & Suess, 2004; Torres *et al*., 2002).

The ROV *Jason* collected snail samples, later used to collect that bacteria used in this genome project, from the northern portion of Hydrate Ridge (44º 40.02687' N 125º 5.99969' W). *Provanna* snails with attached microbial epibionts from the same site as those sequenced here were collected and placed in a tank containing sulfide-rich sediments surrounding active microbial degradation of bone, and were seen to exhibit atypical behavior, wherein they would cross into sulfidic zones, turn onto their shells, and bury the attached sulfide-oxidizing bacteria into the sediment (Flood & Bailey, unpublished). It is currently unknown whether the snails were exhibiting this behavior as a protective mechanism from the toxic sulfide, or exposing their epibionts to electron donors and subsequently feeding on these attached cells. The snails are unable to feed on the bacteria attached to their own shells, but it is possible this is a group-beneficial feeding strategy among snails, as they would be able to graze on the shells other snails present in the sediments. The cross-phyla breadth of the epibiont community sequenced on the surface of *Thiomargarita nelsonii* suggests a complex ecological relationship between the snails,

their bacterial companions, the epibiont community on those companions, and the geochemical surroundings.

4.3. *Thiomargarita* in the Environment

The discovery of large, sulfide-oxidizing organisms in sulfidic marine sediments distributed across the globe illustrates the potential importance of these bacteria in the global sulfur cycle. As several major clades of *Thiotrichaceae* were only recently recognized, these bacteria still lack investigation and understanding of their makeup. The role of endosymbionts and their hosts has recently been characterized by Nakagawa *et al*. (2014) for a snail and its gammaproteobacterial endosymbiont. The relationship in question is one of a snail in a harsh geochemical environment and its bacterial symbionts that aid in its survival in such conditions. Understanding the role of *Beggiatoa* sp. (a close relative of *Thiomargarita*), its physiology, genetics, and how it interacts with its geochemical environment has been evaluated in previous studies (Mußmann *et al*., 2007; Girnth *et al*., 2011; Tang *et al*., 2013), but the connection between *Thiomargarita* and its ecological niche, morphology, and genetics remains unclear.

Bacterial species on the micron scale typically are diffusion limited, that is, the surface area to volume ratio of small bacterial generally dictates their maximum size if they are to optimize the diffusion of electron donors and acceptors into the cell (Schulz & Jørgensen, 2001). However, the disparity in surface-area-to-volume between an average bacterial cell (1 μm) to those observed in the largest sulfur oxidizing bacteria (750 μm) varies by over 3 orders of magnitude, indicating that *Thiomargarita*'s size is directly related to its survival, as it is likely its electron-acceptor storing capabilities are selected for in its ecological niche. In contrast to *Beggiatoa* and *Thioploca*, *T. nelsonii* is not motile (Schulz *et al*., 1999), though a limited rolling motility has been observed in some *Thiomargarita* (Salman *et al*., 2011). Thus, the cells can neither follow the overlapping boundary of sulfide with oxygen or nitrate as has been seen in *Beggiatoa*; nor can they span the geochemical horizons of the sediment like filamentous

Thioploca. The sediments inhabited by *T. nelsonii* are a biologically-diverse ecological community with very high sulfide concentrations of >10 mM (Schulz & Jørgensen, 2001). This sulfide production is powered by some of the highest AOM rates measured in marine sediments (Treude *et al*., 2003). Off the coast of Namibia, the highest density cell density of *Thiomargarita namibiensis* is found near the sediment surface (Schulz *et al*., 1999), but living cells containing nitrate can be found down to >10 cm (Schulz & Jørgensen, 2001). Apparently, these buried *Thiomargarita* cells can only come in contact with nitrate when the loose sediment becomes suspended into the water column. This may happen as a result of turbulence by wave motion, macrofaunal disturbance, or methane bubbling, which is known to occur regularly in areas inhabited by *Thiomargarita* (Copenhagen, 1953). In the Hydrate Ridge environment, the marine geochemical gradients can change in response to seasonal conditions, and in the case of the *Provanna*-attached *Thiomargarita*, they can move in and out of different geochemical conditions. The similarities of the dynamic geochemical conditions observed in these geographically-separated environments selects for organisms share similar ecophysiologies, namely those shared by *Thiomargarita*-like organisms. The substrate-attached morphotype described here is a subset of this comparably-equipped group of organisms; a morphotype that resides in its own ecological niche, likely possessing unique and distinctive adaptions to the environmental pressures it experiences.

Thiomargarita's ability to store large amounts of intracellular nitrate allows the bacterium to endure highly dynamic environmental conditions, provided the presence of an electron acceptor. If one assumes a rate of nitrate reduction per volume of cytoplasm similar to that of *Thioploca* as measured by Otte *et al*. (1999), an average-sized *Thiomargarita* coccoid cell (~250 μm) should be able to respire for at least 40–50 days without taking up new nitrate (Schulz *et al*., 1999), though the observed survival periods of *Thiomargarita* have been measured to be much longer (Schulz & Jørgensen, 2001).

At least some *Thiomargarita* appear to tolerate high oxygen concentrations and can survive exposure to oxygen-saturated waters. *Thiomargarita namibiensis* has also been shown to use oxygen as an electron acceptor for the oxidation of sulfide (Schulz & de Beer, 2002). By measuring radial oxygen microprofiles around single cells of *Thiomargarita*, previous studies have indicated that cells take up oxygen and the uptake of oxygen is greatly enhanced in the presence of sulfide and vice versa. The presence of both high- and low-affinity cytochrome c oxidases strongly indicates a versatile oxygen-utilizing metabolism. This usage of oxygen also suggests another adaptation to resuspension into the water column, an occurrence much more likely in the coccoidal morphotype due to the lack of substrate attachment common in the elongate form. If the highly sulfidic sediments are physically mixed with the oxic water column or oxygen penetrates the sediments, *Thiomargarita namibiensis* has been shown to use the available oxygen to oxidize internal sulfur or sulfide distributed into the water column without relying on the stored intracellular nitrate (Schulz & Jørgensen, 2001). To compare, the *Provanna*-attached *Thiomargarita nelsonii* experiences similarly dynamic geochemical conditions, as attachment to a mobile snail subjects the bacterial species to differing conditions depending on the placement of the gastropod both spatially and temporally, in relation to the oxic-anoxic boundary. While both the sediment-inhabiting *Thiomargarita namibiensis* and the host-attached *Thiomargarita nelsonii* compensate for their lack of active motility with metabolic flexibility, the mobility imparted to *T. nelsonii* by its host affords a greater opportunity for suspension in the water column, in addition to a more constant exposure to electron acceptors.

4.4. Genome and Metabolism Discussion

4.4.1. Nitrate

Under anoxic conditions, *Thiomargarita* has been shown to respire vacuolar nitrate (McHatton *et al*., 1996; Sayama *et al*., 2005; Kamp *et al*., 2006) with concentrations up to 800 mM NO_3^- measured in *Thiomargarita namibiensis* (Schulz, 2006). The annotated draft genome for Bud 1 encodes both membrane-bound nitrate reductases (*narGHIJ*; Tmarg_01615-01618, Tmarg_11764-11767) and periplasmic nitrate reductases (*napAGH*; Tmarg_12325-12327, Tmarg_06472-06478) (Fig. 6A). Mobile genetic elements annotated as group II catalytic introns were found to be spliced in between the *napA* gene and the *napGH* genes (Tmarg_06472-06478), in contrast to the other full, uninterrupted operon (Tmarg_12325-12327). Nitrate reductases can also operate in the reverse direction in nitrite-oxidizing bacteria, where they are considered nitrite oxidoreductases (Starkenburg *et al*., 2006). As there is physiological evidence for nitrite oxidation in anaerobic ammonium oxidizing (anammox) bacteria with *narG* as candidate enzyme (Schulz & Jørgensen, 2001), the use of environmental and intracellular nitrite is as an electron donor is possible in *Thiomargarita*, under nutrient stress. In Beutler *et al*. (2012), it was shown in a similar sulfide-oxidizing bacterium, *Ca. Allobeggiatoa halophila*, that the reduction of oxidized nitrogen species takes place both in the cytoplasm and in the vacuolar membrane. According to their model, nitrate is reduced in the cytoplasm by a nitrate reductase, resulting in a proton flux through the vacuolar membrane, thereby creating a pH gradient over the membrane. This pH gradient is responsible for a proton motive force that not only allows for the synthesis of ATP and pyrophosphates, but also uptake of additional nitrate from the environment. Nitrite reductases are responsible for the further reduction of the nitrite to nitric oxide (NO) inside the vacuole. Additionally, Beutler *et al.* propose that NO reductases mediating NO consumption are present in the vacuolar membrane, and reduce NO to the intermediate N_2O.

The preferred pathway and regulation of nitrate respiration in *Thiomargarita* remains incompletely understood (Jørgensen & Nelson, 2004). It is assumed that the resulting nitrogen species from dissimilatory nitrate reduction through the pathways linked to sulfide oxidation in *T. nelsonii* is ammonia or an ammonium ion (dissimilatory nitrate reduction to ammonia; DNRA)

(Teske & Nelson, 2004). However, *Thiomargarita namibiensis* cells that were removed from the sediment and incubated with ^{15}N-labeled nitrate revealed a build up of ^{15}N-labeled molecular nitrogen (Schulz, 2006), providing evidence for the presence of a full denitrification pathway. Brunet & Garcia-Gil (1996) showed that only in sulfide-enriched slurries was nitrate appreciably reduced to ammonia, while at low concentrations of sulfide, denitrification was preferentially utilized. A correlation of dissimilatory nitrate reduction to ammonia (DNRA) to oxidation of sulfide to elemental sulfur was also observed, a process potentially responsible for the production of the elemental sulfur globules seen in *Thiomargarita* sp. Sulfide inhibition of NO- and N_2O-reductases is proposed as being responsible for the preferential reduction of nitrate to ammonia. Analysis of the *Thiomargarita* genome revealed the presence of an NAD(P)H-dependent nitrite reductase (*nirBD*, the large and small subunits respectively) which is responsible for the direct reduction of nitrite to ammonia. The genes encoding both subunits (Tmarg_08782, Tmarg_08783) were found on the same operon, and were shown to have a best BLASTP match to the Gammaproteobacterium *Thioalkavibrio*, a chemolithoautotrophic sulfur-oxidizing bacterium commonly isolated from soda lake sediments. The intact nature of the two subunits and its close affiliation to a closely-related bacterium indicates the likelihood of a correct annotation and assembly. Further in the denitrification pathway, 3 nitrite reductases (*nirS*; Tmarg_03321, Tmarg_07219, Tmarg_12671) and four nitric oxide reductases (*norB*; Tmarg_03323, Tmarg_08827 and *norC*; Tmarg_08828 and Tmarg_03322) reduce nitrite and nitric oxide, respectively, to nitrous oxide (Fig. 6A). In two of the genes encoding for the *nirS* enzyme, the *norBC* genes are immediately downstream, suggesting linkage of the subsequent transcription. Lastly, the presence of a nitrous oxide reductase (*nosZ*, Tmarg_08173-4) is responsible for the final step in the denitrification pathway, a reduction of N_2O to dinitrogen. This enzyme was predicted in *Beggiatoa* sp. PS (Mußmann *et al*., 2007), but was not observed. In summary, the genomic data presented here provide the first unambiguous genomic evidence of the significant denitrification potential of large marine sulfur bacteria.

4.4.2. Sulfur

Recent studies on nitrate-respiring *Beggiatoa* have pointed to a two-step oxidation of sulfide (Sayama *et al*., 2005; Kamp *et al*., 2006). In the absence of oxygen, sulfide is oxidized to elemental sulfur and sulfate by reduction of intracellular nitrate. The further oxidation of the internally stored sulfur globules likely occurs in the absence of sulfide, using environmental oxygen as the terminal electron acceptor. The initial oxidation of hydrogen sulfide to elemental sulfur is likely catalyzed via one of two pathways: (1) a sulfide quinone oxidoreductase (*sqr*; Bud 2) or (2) a flavocytochrome c/sulfide dehydrogenase (*fccAB*; Bud 2) (Fig. 6B). The two genes seem to have differing expression at varying concentrations of sulfide, with Sqr preferentially oxidizing under high sulfide, low oxygen conditions and FccAB more prevalent at low sulfide, high oxygen concentrations in the upper sediment (Reinartz *et al*., 1998; Eddie & Hanson, 2013).

In *T. nelsonii* and related sulfide-oxidizing bacteria, the reverse dissimilatory sulfite reduction (rDSR) pathway is likely responsible for the continued oxidation of intracellular elemental sulfur to sulfite (Dahl *et al*., 2005). The genes encoding the rDSR pathway include the common dissimilatory sulfite reductase genes (*dsrABC*; Bud 2) and the complete transmembrane electron-transporting complex (*dsrKMJ*; Tmarg_08388, Tmarg_08387, Tmarg_08389; *dsrOP*; Bud 2). The secondary DsrKMJOP complex is present in three separate fragments throughout the *T. nelsonii* genome, with the former three genes encoding a protein with high similarity to heterosulfide reductase from methanogenic archaea, a triheme cytochrome c mediating transport across the cytoplasm, and a membrane-bound b-type cytochrome, respectively. Their closest phylogenetically-affiliated homologs, as determined using the IMG database, are a conserved Fe-S oxioreductase, a respiratory nitrate reductase, and an electron transport protein. The *dsrOP* encoding genes are absent from the Bud 1 sample, but have multiple strong BLASTP matches in the Bud 2 dataset (*dsrO*, 4e-117; *dsrP*, 0.0), and respectively encode a periplasmic iron-sulfur protein and integral membrane protein (Dahl *et al*., 2005; Ulrich *et al*., 2013). The rDSR pathway

as described here is likely essential for mobilizing intracellular elemental sulfur and binding the produced sulfite for the final oxidation step to sulfate.

The oxidation of sulfur species to sulfate is the final step of the sulfur oxidation pathway, and can be reached in three ways. First, the sulfite produced from the rDSR pathway can be oxidized and phosphorylated via an adenosin-phosphosulfate (APS) reductase (*aprAB*; Tmarg_02567-02568) to produce APS. The APS is further dephosphorylated to produce an ATP and a sulfate ion via an ATP sulfurylase (Tmarg_05810, Tmarg_00628). It has been proposed by Mußmann *et al*. (2007) and Haveman *et al*. (2004) that heterodisulfide reductases (*hdrABC*; Tmarg_02860-02862) transport electrons to the AprAB enzymes, a role previously described in sulfate-reducing bacteria. The second potential method of producing sulfate utilizes sulfite dehydrogenase (*sorA*; Tmarg_09133) to further oxidize the sulfite produced from the rDSR pathway by employing ferricytochromes as electron acceptors. The flexibility of the sulfite-sulfate oxidation pathway further increases options available for *Thiomargarita* to metabolize the available environmental and intracellular sulfur species (Mußmann *et al*., 2007). The final metabolic enzymes found in *T. nelsonii* capable of producing sulfate complete the sulfur oxidation pathway (*soxABX*, Bud 2; *soxYZ*, Tmarg_02753-4, Tmarg_06834-5), which is capable of both oxidizing sulfite and thiosulfate to ultimately produce sulfate.

A complete SOX system, as found in *Paracoccus pantotrophus* (Friedrich *et al*., 2001), is capable of fully oxidizing sulfite and thiosulfate to produce sulfate without producing intermediates like elemental sulfur. The incomplete SOX system present in *Thiomargarita* is thought to work in a co-dependent fashion, with SoxXA initiating the cycle by aiding in the bonding of thiosulfate with the SoxYZ group to form a SoxY-thiocysteine-S-sulfate complex (Fig. 7A). SoxB would hydrolyze sulfate from the thiocysteine-S-sulfate residue to give thiocysteine-S (Fig. 7B). The method of hydrolyzing the final sulfur from the thiocysteine to produce elemental sulfur is not wholly understood (Fig. 7C), though the presence of a SoxCD

group (not found) has been shown to form an additional sulfate in place of the final sulfur (Haveman *et al*., 2004). Therefore, it is likely that an unknown process is responsible for the production of the intracellular sulfur noted in *Thiomargarita*, or that the SoxCD group is not represented in our incomplete genome. It is also possible that the cytochrome c oxidase just upstream of the *soxYZ* genes plays some role (Tmarg_02752). Fig. 7D illustrates the similar mechanisms by which the SOX system would carry out oxidation to sulfate via oxidation of sulfite, as opposed to thiosulfate, with the absence of a step to create elemental sulfur.

4.4.3. Phosphorous

Polyphosphate metabolism occurs in most organisms, with prokaryotes and some fungi able to store large amounts intracellularly (Kornberg, 1995). It has been shown that *Thiomargarita* and related large sulfur bacteria are capable of creating large polyphosphate inclusions to utilize them under anoxic or sulfidic conditions for sulfide oxidation (Schulz & Schulz, 2005; Goldhammer *et al*., 2010; Brock & Schulz, 2011). In oxic conditions, it is thought that *Thiomargarita* accumulates environmental nitrate and phosphate in the cytoplasm, in order to readily adapt to changing geochemical conditions. While this chemical uptake and release has been experimentally verified, a mechanism for synthesis and hydrolysis of polyphosphate has yet to be shown in *Thiomargarita*. Analysis of the *T. nelsonii* datasets reveal the presence of genes that encode for polyphosphate kinases (Tmarg_11806, Tmarg_02749, Tmarg_08103), the latter two directly downstream from a putative transposase (Tmarg_08101, Tmarg_02747). These kinases are likely responsible for the formation of polyphosphate chains from ATP as shown in Fig. 6C. The hydrolysis of these molecules is likely carried out by polyphosphate pyrophosphohydrolases, which are encoded for multiple times in the *T. nelsonii* genome (Tmarg_08051, Tmarg_11732, Tmarg_04433, Tmarg_07574). These genes are also annotated as intermediates in purine metabolic pathways, responsible for the release of phosphate from Guanosine-5'-triphosphate (GTP). The release of phosphate into the environment from large, vacuolated sulfur bacteria has been shown in prior experiments (Schulz & Schulz, 2005; Brock &

Schulz, 2011), and it was likely through this mechanism that the phosphate was taken into the cell, and subsequently hydrolyzed. The polyphosphate stored intracellularly has been implicated in numerous studies for geologic deposition in phosphatic minerals and phosphorites (Williams & Reimers., 1983; Reimers *et al*., 1990; Nathan *et al*., 1993; Krajewski *et al*., 1994; Goldhammer *et al*. 2010; Crosby & Bailey, 2012), and has potential as a large phosphorous sink for environmental phosphorous. Diagenetic precipitation of phosphatic minerals has been reported from the Santa Barbara Basin, wherein elevated pore water nitrate concentrations were correlated to the metabolic activity of large sulfur bacteria (Reimers *et al*., 1996). Schulz & Schulz (2005) were also able to show that under anoxic conditions, *Thiomargarita namibiensis* released enough phosphate to account for the formation of hydroxyapatite observed to be precipitating in modern sediments off the coast of Namibia. Additionally, Goldhammer *et al*. (2010) incubated *Thiomargarita*- and *Beggiatoa*-rich sediments with a ^{33}P radiotracer to determine the ultimate fate of pore-water phosphorous. Under both oxic and anoxic conditions, the bacteria accumulated ^{33}P intracellularly, and catalyzed the formation of phosphate to apatite, with the largest amount of apatite formation under anoxia. The removal of this phosphorous from the water column via apatite sequestration (and ultimately phosphorite deposition) is likely a long-term geologic sink for phosphorous. The occurrence of modern or recent phosphorites and the high *Beggiatoaceae* biomass are roughly correlated geographically, though phosphorites are not known from methane seep or hydrothermal vent settings. As phosphorite precipitation through the hydrolysis and release of phosphate could be a relatively common phenomenon in marine sediment surface, it could have also preserved potential microbial presence in past phosphatic minerals (e.g. Bailey *et al*., 2013).

4.4.4. Oxygen

In order to survive changing geochemical conditions, it is necessary for *Thiomargarita* to be able to make use of multiple electron acceptors, including oxygen. Intracellular nitrate is thought to serve as an energetically favorable acceptor when the bacteria reside in anoxic

sediments, but the respiration of oxygen is also important, given the likely transportation of the attached cells between the sulfidic, anoxic sediments (Fig. 8A) and the oxic water column via movement from the host snails (Fig. 8B). It has been shown by Girnth *et al*. (2011) that highly dynamic geochemical conditions favor *Thiomargarita* species over other sulfur-oxidizing bacteria, illustrating a potential survival strategy for bacteria both associated with mobile hosts and highly variable environmental chemistries. In the organic-rich sediments, oxygen penetrates only the upper few millimeters beneath the sediment-water interface. The chemotactic response away from high oxygen concentrations that has been observed in other large, vacuolated sulfur bacteria (Mußmann *et al*., 2007) has not been observed in *Thiomargarita*, but it is likely that the oxidation of sulfur compounds is most energetically favorable in microoxic conditions. The presence of the low-affinity terminal oxidase (cytochrome c bb_3) was noted in multiple locations in the genome (Tmarg_00823-00825, Tmarg_12758-12760, Tmarg_08520-08522), and is assumed to be used under microoxic concentrations, and is likely the chief terminal oxidase utilized. The high-affinity terminal oxidase (cytochrome c aa3) was not present in the Bud 1 dataset, but a putative cytochrome c aa3-oxidase is present in the Bud 2 dataset, and is likely responsible for oxygen utilization under high oxygen concentrations. Differential expression of cytochrome c oxidases has been shown in freshwater *Beggiatoa* species (Muntian *et al*., 2005), and is likely occurring in *Thiomargarita* under highly variable oxygen concentrations.

4.4.5. Carbon Metabolism and Heterotrophy

The flux of hydrocarbons from subsurface methane clathrates is an important source of carbon to microbial communities at Hydrate Ridge and other cold seep sites. The subsequent bacterial oxidation of this abundant organic carbon in turn has an important impact on the seep environment. Bacterial oxidation of hydrocarbons lowers the concentration of oxygen in the sediment and makes conditions favorable for the bacterial reduction of sulfate (Aharon & Fu, 2000). The sediment at methane seeps would therefore contain an ample supply of H_2S, in

addition to organic compounds (hydrocarbons and metabolic intermediates derived from heterotrophic consumption of those hydrocarbons). These conditions would support microbial heterotrophy as well as autotrophy. Chemoautotrophic growth in the closely-related bacteria Beggiatoa sp. has been described by Nelson *et al*. (1986) in oxygen-sulfide microgradients. Analyses done both on metabolic enzymes and carbon isotope analysis indicate that *Beggiatoa* sp. preferentially use CO_2 generated by the bacterial oxidation of hydrocarbons as their carbon source, and oxidize H_2S for energy (Nelson *et al*., 1989; Sassen *et al.*, 1993).

In *T. nelsonii,* fixation of CO_2 for autotrophic growth is facilitated by a form-I ribulose-bisphosphate carboxylase oxygenase (RuBisCO, Tmarg_03576, Tmarg_05222). Additionally, as has been shown in *Beggiatoa* sp., phosphoribulokinase (Tmarg_08893) and carbonic anhydrase (Tmarg_07234, Tmarg_09509) genes were found. However, closely related *Beggiatoa* species (*B. alba*, *B. leptomitiformis*, *Beggiatoa* sp.) have been shown to grow heterotrophically on acetate and other organic carbon sources (Faust & Wolfe, 1961; Strohl *et al*., 1981; Grabovich *et al*., 1998). The *Thiomargarita* genome possesses acetate-cation symporters (Tmarg_03415, Tmarg_06685) to shuttle acetate into the cell. *Thiomargarita* has additionally been shown to stabilize sulfur oxidation gradients in the presence of low concentrations of acetate in the environment (Schulz & de Beer, 2002). As has been observed in *Beggiatoa*, during growth on acetate or organic carbon, the glyoxylate cycle is likely used for gluconeogenesis (Strohl *et al*., 1981; Muntian *et al*., 2005). As has been shown in the *Beggiatoa* sp. genome (Mußmann *et al*., 2007), key enzymes in the tricarboxylic acid (TCA) cycle—malate synthase and isocitrate lyase —were not identified in the sequenced genome. However, due to the presence of several key enzymes (isocitrate, succinate, malate, and 2-oxoglutarate dehydrogenase, fumarate dehydratase, and succinyl-coenzyme A synthase; Tmarg_04553, Tmarg_08994-6, Tmarg_09218, Tmarg_05369, Tmarg_13445, Tmarg_06689, Tmarg_10119-20) and the consistency with previous findings, it is possible that *Thiomargarita* possesses an incomplete TCA cycle, though the absence of these key enzymes can also be attributed to the incomplete genome. Additionally,

the presence of glycolate oxidase subunits (*glcDEF*, Tmarg_04834-6, Tmarg_10672-4) suggests glycolate utilization by *Thiomargarita*, as has been seen in *Beggiatoa* sp. (Mußmann *et al*., 2007). Photosynthetic organisms and some photoautotrophic chemolithotrophs create glycolate, and while the former organisms are thought to co-occur with the order *Thiotrichales*, it is possible that these glycolate utilization genes are remnants of horizontal transfer similar to that identified elsewhere in the genome.

The synthesis of glycogen or polyglucoses in *Thiomargarita* has been previously recorded (Schulz & Schulz, 2005), and both genome datasets reveal the capability to synthesize glycogen preferentially under oxic conditions, as shown by genes encoding glycogen synthase (Tmarg_04661, Tmarg_06659, Tmarg_06941) and glycogen-debranching enzymes (glucosyltransferase and beta-glucosidase; Tmarg_07642-3. Tmarg_06849; Tmarg_00898). *T. nelsonii* could also produce ATP via phosphorylation from pyruvate by way of a potential fermentative lactate dehydrogenase (*ldh*, Tmarg_00514). This fermentation of storage compounds may enable *Thiomargarita* to survive during periods of oxygen, and nitrate depletion, for example, when the oxic–anoxic interface is located above the sediment-water interface (Fig. 8A).

4.4.6. Cell Division and Shape

Cell wall biogenesis proteins, crucial for the formation of the laterally-extensive cell wall observed in large, vacuolated sulfur bacteria, were putatively transferred in some amount from cyanobacteria and/or Alpha-, Beta-, and Deltaproteobacteria to *Thiomargarita* based on best match comparisons to previously sequenced organisms. It is also possible that the dimorphic morphology observed in *Thiomargarita nelsonii* is the result of a horizontal gene transfer event, as similar elongate and budding morphologies have been observed in cyanobacterial species (*Chamaesiphon*) and Alphaproteobacteria (*Caulobacter crescentus*) (Brun & Janakiraman, 2000). The budding of the coccoidal *Thiomargarita* cells is likely a response to nutrient stress, as elongation increases the surface area to volume ratio from the original spherical cell (i.e. a rod has

a more surface area than a sphere of the same volume) (Young, 2006). Reductive division has been seen in the Costa Rica Margin *Thiomargarita* sp. inhabiting sediments near the budding *Thiomargarita* morphotype (Kalentra *et al*., 2005; Bailey *et al*., 2007). It is feasible that the *Thiomargarita* sp. observed belong to the same species, yet exist in different stages of a complex life cycle. One possibility is that elongation increases surface area in attached cells that have access to nitrate and other metabolites can readily enlarge or divide; whereas reductive division parcels stored nitrate into smaller, more easily dispersible cells with higher surface area to volume ratios when nitrate is limited. The greater the radius of a spherical *Thiomargarita* cell, the larger its vacuole, and thus its potential store of nitrate, but at the cost of decreasing its surface area-to-volume ratio. Growth of a cell into an elongate filament, as opposed to simply expanding the sphere volume, minimizes the decrease in surface area-to-volume ratio that accompanies growth (Bailey *et al*., 2011).

5. Conclusion

This study uses a combination of whole genome amplification, Illumina paired-end sequencing, and binning via ESOM to sequence the genome of *Thiomargarita nelsonii*, an organism that has thus far proven unculturable. *Thiomargarita* possesses the genes encoding the denitrification pathway that reduces nitrate to dinitrogen, the dissimilatory nitrate reduction to ammonia (DNRA) pathway to reduce nitrate to ammonia, and the genes for the synthesis and hydrolysis of polyphosphates. In the absence of an isolated culture, these genes are our only lines of evidence that these transformations are catalyzed by *Thiomargarita nelsonii*. It is also likely that *Thiomargarita* has also received some of its genetic material from other prokaryotic species via horizontal gene transfer, rather than from vertical descent with modification. The role that *Thiomargarita* and other similarly-equipped large sulfur-oxidizing bacteria play in the global phosphorous, sulfur, and nitrogen cycles is increasing in importance with further study. The sequestration of pore water phosphorous into phosphatic minerals and phosphorites is mediated by the polyphosphate accumulation and release by *Thiomargarita* sp. and related organisms. The

cycling of toxic, reduced sulfur species is crucial to metabolism of this bacterial species and the long-term viability of the ecological niche it inhabits. The denitrification potential and the DNRA exhibited by *Thiomargarita* has been shown in a laboratory environment, but prior to this study, had never been shown in sequence data from this clade. The metabolic activity and availability of electron donors and acceptors seems to also inform the morphotype of the bacteria. In times of nutrient abundance, an elongate form can be taken to increase surface area, and thusly uptake of the oxygen, nitrate, and phosphorous to be stored intracellularly. The still poorly understood adaptation of the non-motile *Thiomargarita* to environments with dynamic geochemical conditions and redox gradients illustrates the need to fully explore all lines of investigation available in the absence of pure culturing. Indeed, the sequencing and analysis of this genome and others yet to come, may help improve our chances of growing and studying these organisms in the laboratory.

6. Bibliography

Aharon P, Fu B (2000) Microbial sulfate reduction rates and sulfur and oxygen isotope fractionations at oil and gas seeps in deepwater Gulf of Mexico. *Geochimica et Cosmochimica Acta* **64**, 233-246.

Andrews S (2010) FASTQC A quality control tool for high throughput sequence data URL http://wwwbioinformaticsbabrahamacuk/projects/fastqc

Aronesty E (2011) Command-line tools for processing biological sequencing data *Expression analysis.*

Bailey JV, Joye SB, Kalanetra KM, Flood BE, Corsetti FA (2006) Evidence of giant sulphur bacteria in Neoproterozoic phosphorites. *Nature* **445**, 198-201.

Bailey JV, Salman V, Rouse GW, Schulz-Vogt HN, Levin LA, & Orphan VJ (2011) Dimorphism in methane seep-dwelling ecotypes of the largest known bacteria. *The ISME journal* **5***,* 1926-1935.

Bell LC, Richardson DJ, Ferguson SJ (1990) Periplasmic and membrane-bound respiratory nitrate reductases in *Thiosphaera pantotropha*: The periplasmic enzyme catalyzes the first step in aerobic denitrification. *FEBS letters* **265**, 85-87.

Beller HR, Chain PS, Letain TE, Chakicherla A, Larimer FW, Richardson PM, Kelly DP (2006) The genome sequence of the obligately chemolithoautotrophic, facultatively anaerobic bacterium *Thiobacillus denitrificans. Journal of bacteriology* **188**,1473-1488.

Besiktepe S, Tang K, Vila M, Simó R (2004) Dimethylated sulfur compounds in seawater, seston and mesozooplankton in the seas around Turkey. *Deep Sea Research Part I: Oceanographic Research Papers* **51**, 1179-1197.

Beutler M, Milucka J, Hinck S, Schreiber F, Brock J, Mußmann M, Schulz-Vogt H N, de Beer D (2012) Vacuolar respiration of nitrate coupled to energy conservation in filamentous *Beggiatoaceae*. *Environmental Microbiology* **14**, 2911–2919.

Binga EK, Lasken RS, Neufeld JD (2008) Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *The ISME Journal* **2**, 233-241.

Blankenberg D, Kuster GV, Coraor N, Ananda G, Lazarus R, Mangan M, Taylor J (2010) Galaxy: A Web Based Genome Analysis Tool for Experimentalists. *Current protocols in molecular biology* **19**, 1-21.

Boetius A, Ravenschlag K, Schubert CJ, Rickert D, Widdel F, Gieseke A, Pfannkuche O (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane *Nature* **407**, 623-626.

Boetius A, Suess E (2004) Hydrate Ridge: a natural laboratory for the study of microbial life fueled by methane from near-surface gas hydrates. *Chemical Geology* **205**, 291-310.

Bowles M, Joye S (2010). High rates of denitrification and nitrate removal in cold seep sediments. *The ISME journal* **5**, 565–567

Brimblecombe P, Shooter D (1986) Photo-oxidation of dimethylsulphide in aqueous solution. *Marine Chemistry* **19**, 343-353.

Brock J, Schulz-Vogt H (2011) Sulfide induces phosphate release from polyphosphate in cultures of a marine *Beggiatoa* strain. *The ISME journal* **5**, 497-506.

Brun YV, Janakiraman R (2000) The dimorphic life cycle of Caulobacter and stalked bacteria. In: Prokaryotic Development (Brun YV and Shimkets LJ (eds)). *American Society for Microbiology,* 297–317.

Brunet RC, Garcia-Gil LJ (1996) Sulfide-induced dissimilatory nitrate reduction to ammonia in anaerobic freshwater sediments. *FEMS Microbiology Ecology* **21**, 131-138.

Copenhagen W (1953) The periodic mortality of fish in the Walvis region: A phenomenon within the Benguela Current. *Division of Sea Fisheries*.

Crosby CH, Bailey JV (2012) The role of microbes in the formation of modern and ancient phosphatic mineral deposits. *Front. Microbio.* **3**, 241.

Dahl C, Engels S, Pott-Sperling AS, Schulte A, Sander J, Lübbe Y, Brune DC (2005) Novel genes of the *dsr* gene cluster and evidence for close interaction of Dsr proteins during sulfur oxidation in the phototrophic sulfur bacterium *Allochromatium vinosum*. *Journal of Bacteriology* **187**, 1392-1404.

Dichosa AEK, Fitzsimons MS, Lo CC, Weston LL, Preteska LG, *et al* (2012) Artificial Polyploidy Improves Bacterial Single Cell Genome Recovery. *PLoS ONE* **7**.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biology* **10**, R85.

Eddie BJ, Hanson TE (2013) *Chrolobaculum tepidum* displays a complex transcriptional response to sulfide addition. *Journal of Bacteriology* **195**, 399-408

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput *Nucleic Acids Research* **32**, 1792-1797

Faust L, Wolfe R (1961) Enrichment and cultivation of *Beggiatoa alba*. *Journal of Bacteriology* **81**, 99.

Flood BE, Bailey JV, Biddle JF (2014) Horizontal gene transfer and the rock record: comparative genomics of phylogenetically distant bacteria that induce wrinkle structure formation in modern sediments. *Geobiology* doi: 10.1111/gbi.12072

Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J (2001) Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? *Applied and Environmental Microbiology* **67**, 2873-2882.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Taylor J (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Research***15**, 1451-1455.

Girnth AC, Grünke S, Lichtschlag A, Felden J, Knittel K, Wenzhöfel,F, Boetius A (2011) A novel, mat-forming *Thiomargarita* population associated with a sulfidic fluid flow from a deep-sea mud volcano. *Environmental Microbiology* **13,** 495-505.

Goecks J, Nekrutenko A, Taylor J, Team TG (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **11**, R86.

Goldhammer T, Brüchert V, Ferdelman TG, Zabel M (2010) Microbial sequestration of phosphorus in anoxic upwelling sediments. *Nature Geoscience* **3**, 557-561.

Gonzalez JM, Portillo MC, Saiz-Jimenez C (2005) Multiple displacement amplification as a pre-polymerase chain reaction (pre-PCR) to process difficult to amplify samples and low copy number sequences from natural environments. *Environmental Microbiology* **7**, 1024-1028.

Grabovich MY, Dubinina G, Lebedeva VY, Churikova V (1998) Mixotrophic and lithoheterotrophic growth of the freshwater filamentous sulfur bacterium *Beggiatoa leptomitiformis* D-402. *Microbiology* **67**, 383-388.

Green-Saxena A, Dekas AE, Dalleska VF, Orphan VJ (2013) Nitrate-based niche differentiation by distinct sulfate-reducing bacteria involved in the anaerobic oxidation of methane. *The ISME journal* **8**, 150-163.

Grünke S, Felden J, Lichtschlag A, Girnth AC, De Beer D, Wenzhöfer F, Boetius A (2011) Niche differentiation among mat-forming, sulfide-oxidizing bacteria at cold seeps of the Nile Deep Sea Fan (Eastern Mediterranean Sea). *Geobiology* **9**, 330-348.

Hagen KD, Nelson DC (1997) Use of reduced sulfur compounds by *Beggiatoa* spp: Enzymology and physiology of marine and freshwater strains in homogeneous and gradient cultures. *Applied and Environmental Microbiology* **63**, 3957-3964.

Haveman SA, Greene EA, Stilwell CP, Voordouw JK, Voordouw G (2004) Physiological and gene expression analysis of inhibition of *Desulfovibrio vulgaris Hildenborough* by nitrite. *Journal of Bacteriology* **186**, 7944-7950.

Hensen D, Sperling D, Trüper HG, Brune DC, Dahl C (2006) Thiosulphate oxidation in the phototrophic sulphur bacterium *Allochromatium vinosum*. *Molecular Microbiology* **62,** 794-810.

Hinsley AP, Berks BC (2002) Specificity of respiratory pathways involved in the reduction of sulfur compounds by *Salmonella enterica. Microbiology* **148**, 3631-3638.
Hipp WM, Pott AS, Thum-Schmitz N, Faath I, Dahl C, Trüper HG (1997) Towards the phylogeny of APS reductases and sirohaem sulfite reductases in sulfate-reducing and sulfur-oxidizing prokaryotes. *Microbiology* **143**, 2891-2902.
Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research* **9**, 868-877.
Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Research* **21**, 1552-1560.
Jørgensen BB, Nelson DC (2004) Sulfide oxidation in marine sediments: Geochemistry meets microbiology In: *Sulfur biogeochemistry—Past and present Boulder* (Amend JP, Edwards KJ, Lyons TW) *Geological Society of America,* 63–81.
Joye SB, Boetius A, Orcutt BN, Montoya JP, Schulz HN, Erickson MJ, Lugo SK (2004) The anaerobic oxidation of methane and sulfate reduction in sediments from Gulf of Mexico cold seeps. *Chemical Geology* **205**, 219-238.
Kalanetra K, Joye S, Sunseri N, Nelson D (2005) Novel vacuolate sulfur bacteria from the Gulf of Mexico reproduce by reductive division in three dimensions. *Environmental Microbiology* **7**, 1451-1460.
Kamp A, Stief P, Schulz-Vogt HN (2006) Anaerobic sulfide oxidation with nitrate by a freshwater *Beggiatoa* enrichment culture. *Applied and Environmental Microbiology* **72**, 4755-4760.
Karl DM (2014) Microbially mediated transformations of phosphorous in the sea: new views of an old cycle. *The Annual Review of Marine Science* **6**, 279-337.
Knittel K, Boetius A, Lemke A, Eilers H, Lochte K, Pfannkuche O et al. (2003) Activity, distribution, and diversity of sulfate reducers and other bacteria in sediments above gas hydrate (Cascadia Margin, Oregon). *Geomicrobiol J* **20**, 269–294.
Kornberg A (1995) Inorganic polyphosphate: toward making a forgotten polymer unforgettable. *Journal of Bacteriology* **177**, 491-496.
Krajewski K, Vancappellen P, Trichet J, Kuhn O, Lucas J, Martinalgarra A, Knight R (1994) Biological processes and apatite formation in sedimentary environments. *Eclogae Geologicae Helvetiae* **87**, 701-745.
Kvist T, Ahring BK, Lasken RS, Westermann P (2007) Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Applied Microbiology and Biotechnology* **74**, 926-935.
Lander ES, Waterman MS (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis *Genomics* **2**, 231-239.
Lasken RS, Stockwell TB (2005) Multiple displacement amplification from single bacterial cells in: *Whole genome amplification* (Hughes S, Lasken RS, editors) *Scion Publishing,* 117–148.
MacGregor B, Biddle J, Harbort C, Matthysse A, Teske A (2013) Sulfide oxidation, nitrate respiration, carbon acquisition, and electron transport pathways suggested by the draft genome of a single orange Guaymas Basin *Beggiatoa* (*Cand Maribeggiatoa*) sp filament. *Marine Genomics* **11**, 53-65.
Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Williams P (2012) IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research* **40**, D115-D122.
McHatton SC, Barry JP, Jannasch HW, Nelson DC (1996) High nitrate concentrations in vacuolate, autotrophic marine *Beggiatoa* spp *Applied and Environmental Microbiology* **62**, 954-958.
Mendell JE, Clements KD, Choat JH, Angert ER (2008) Extreme polyploidy in a large bacterium. *Proceedings of the National Academy of Sciences* **105**, 6730-6734,.
Møller, MM, Nielsen, LP, Jørgensen BB (1985) Oxygen responses and mat formation by *Beggiatoa* spp *Applied and Environmental Microbiology* **50**, 373-382.

Muntian, M, Grabovich, M, Patritskaia, V, & Dubinina, G (2005) Regulation of metabolic and electron transport pathways in the freshwater bacterium *Beggiatoa leptomitiformis* D-402 *Mikrobiologiia* **74**, 452.

Mußmann M, Richter M, Lombardot T, Meyerdierks A, Kuever J, Kube M, Amann R (2005) Clustered genes related to sulfate respiration in uncultured prokaryotes support the theory of their concomitant horizontal transfer. *Journal of bacteriology* **187**, 7126-7137.

Mußmann, M, Hu, F Z, Richter, M, de Beer, D, Preisler, A, Jørgensen, BB, Koopman, WJ (2007) Insights into the genome of large sulfur bacteria revealed by analysis of single filaments *PLoS biology* **5**, 230.

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

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012) MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research* **40**, 155.

Nathan Y, Bremner JM, Loewenthal RE, Monteiro P (1993) Role of bacteria in phosphorite genesis. *Geomicrobiology Journal* **11**, 69-76.

Nauhaus K, Boetius A, Krüger M, Widdel F (2002) In vitro demonstration of anaerobic oxidation of methane coupled to sulphate reduction in sediment from a marine gas hydrate area. *Environmental Microbiology* **4**, 296-305.

Nelson DC, Jannasch HW (1983) Chemoautotrophic growth of a marine *Beggiatoa* in sulfide-gradient cultures. *Archives of Microbiology* **136**, 262-269.

Nelson DC, Revsbech NP, Jorgenson BB (1986) Microoxic- anoxic niche of *Beggiatoa* spp.: microelectrode survey of marine and freshwater strains. *Applied and Environmental Microbiology* **52**, 161-168.

Nelson DC, Williams CA, Farah BA, Shively JM (1989) Occurrence and regulation of calvin cycle enzymes in non-autotrophic Beggiatoa strains. *Archives of Microbiology* **151**, 15-19.

Otte S, Kuenen JG, Nielsen LP, Paerl HW, Zopfi J, Schulz HN, Jørgensen BB (1999) Nitrogen, Carbon, and Sulfur Metabolism in Natural *Thioploca* Samples. *Applied and Environmental Microbiology,* **65**, 3148-3157.

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

Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, *et al* (2007) Targeted access to the genomes of low abundance organisms in complex microbial communities. *Appl. Environ. Microbiol.* **73**, 3205–3214.

Pott AS, Dahl C (1998) Sirohaem sulfite reductase and other proteins encoded by genes at the *dsr* locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur. *Microbiology* **144**, 1881-1894.

Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P (2007) Prediction of effective genome size in metagenomic samples. *Genome biology* **8**, R10.

Raghunathan A, Ferguson HR, Bornarth CJ, Song WM, Driscoll M, *et al* (2005) Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* **71**, 3342–3347.

Reimers C, Kastner M, Garrison R (1990) The role of bacterial mats in phosphate mineralization with particular reference to the Monterey Formation. *Phosphate Deposits of the World* **3**, 300-311.

Reimers C, Ruttenberg KC, Canfield DE, Christiansen MB, Martin JB (1996) Porewater pH and authigenic phases formed in the uppermost sediments of the Santa Barbara Basin. *Geochimica et Cosmochimica Acta* **60**, 4037-4057.

Reinartz M, Tschäpe J, Brüser T, Trüper HG, Dahl C (1998) Sulfide oxidation in the phototrophic sulfur bacterium *Chromatium vinosum. Archives of Microbiology* **170**, 59-68.

Sahling H, Rickert D, Lee RW, Linke P, Suess E (2002) Macrofaunal community structure and sulfide flux at gas hydrate deposits from the Cascadia convergent margin, NE Pacific. *Marine Ecology Progress Series* **231**, 121-138.
Salman V, Amann R, Girnth A-C, Polerecky L, Bailey JV, Høgslund S, Schulz-Vogt H N (2011) A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. *Systematic and Applied Microbiology* **34**, 243-259.
Salman V, Bailey JV, Teske A (2013) Phylogenetic and morphologic complexity of giant sulphur bacteria. *Antonie van Leeuwenhoek* **104**, 169-186.
Sassen R, Roberts HH, Aharon P, Larkin J, Chinn EW, Carney R (1993) Chemosynthetic bacterial mats at cold hydro- carbon seeps, Gulf of Mexico continental slope. *Organic Geochemistry* **20**,77-89.
Sayama M, Risgaard-Petersen N, Nielsen LP, Fossing H, Christensen PB (2005) Impact of bacterial NO_3^- transport on sediment biogeochemistry. *Applied and Environmental Microbiology* **71**, 7575-7577.
Schatz MC, Phillippy AM, Sommer DD, Delcher AL, Puiu D, Narzisi G, Pop M (2013) Hawkeye and AMOS: visualizing and assessing the quality of genome assemblies. *Briefings in Bioinformatics* **14**, 213-224.
Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863-864.
Schulz HN, Brinkhoff T, Ferdelman T, Mariné M, Teske A, Jorgensen B (1999) Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* **284**, 493-495.
Schulz HN, Jørgensen BB (2001) Big bacteria. *Annual Reviews in Microbiology* **55**, 105-137.
Schulz HN, de Beer D (2002) Uptake rates of oxygen and sulfide measured with individual *Thiomargarita namibiensis* cells by using microelectrodes. *Applied and Environmental Microbiology* **68**, 5746-5749.
Schulz HN, Schulz HD (2005) Large sulfur bacteria and the formation of phosphorite. *Science* **307**, 416-418.
Schulz, HN (2006) The genus *Thiomargarita. Prokaryotes*, 1156-1163.
Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol İ (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research* **19**, 1117-1123.
Simpson JT, Durbin R (2010) Efficient construction of an assembly string graph using the FM-index *Bioinformatics* **26**, i367-i373.
Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K (2006) Whole-genome multiple displacement amplification from single cells. *Nature Protocols* **1**, 1965-1970.
Sommer D, Delcher A, Salzberg S, Pop M (2007) Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* **8**, 64.
Starkenburg SR, Chain PS, Sayavedra-Soto LA, Hauser L, Land ML, Larimer FW, Arp D J (2006) Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255. *Applied and Environmental Microbiology* **72**, 2050-2063.
Strohl WR, Cannon GC, Shively JM, Güde H, Hook LA, Lane CM, Larkin JM (1981) Heterotrophic carbon metabolism by *Beggiatoa alba. Journal of Bacteriology* **148**, 572-583.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28,** 2731-2739.
Tang K, Liu K, Jiao N, Zhang Y, Chen C-T A (2013) Functional metagenomic investigations of microbial communities in a shallow-sea hydrothermal system. *PloS one.*
Teske A, Nelson D C (2004) The genera Beggiatoa and Thioploca. In:The prokaryotes: An evolving electronic resource for the microbial community (Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors) *Fischer Verlag.*

Teske A, Sørensen KB (2007) Uncultured archaea in deep marine subsurface sediments: have we caught them all? *The ISME journal* **2**, 3-18.
Torres M, McManus J, Hammond D, De Angelis M, Heeschen K, Colbert S, Suess E (2002) Fluid and chemical fluxes in and out of sediments hosting methane hydrate deposits on Hydrate Ridge, OR, I: Hydrological provinces. *Earth and Planetary Science Letters* **201**, 525-540.
Treude T, Boetius A, Knittel K, Wallmann K, Jørgensen BB (2003) Anaerobic oxidation of methane above gas hydrates at Hydrate Ridge, NE Pacific Ocean. *Marine Ecology Progress Series* **264**, 1-14.
Tryon M, Brown K, Torres M (2002) Fluid and chemical flux in and out of sediments hosting methane hydrate deposits on Hydrate Ridge, OR, II: Hydrological processes. *Earth and Planetary Science Letters* **201**, 541-557.
Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, *et al* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43.
Ulrich T, Lanzén A, Stokke R, Pedersen RB, Bayer C, Thorset IH, Schleper C, Steen IH, Øvreas L (2013) Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated meta-omics. *Environmental Microbiology* doi:10.1111/1462-2920.12283
Ultsch A, Mörchen F (2005) ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM.
Viswanathan V K (2012) Sizing up microbes. *Gut microbes* **3**, 483-484.
Waldron F (1900) On the appearance and disappearance of a mud island at Walfish Bay. *Transactions of the South African Philosophical Society* **11**, 185-188.
Wang H, Tseng CP, Gunsalus RP (1999) The napF and narG nitrate reductase operons in *Escherichia coli* are differentially expressed in response to submicromolar concentrations of nitrate but not nitrite. *Journal of Bacteriology* **181**, 5303-5308.
Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991) 16S ribosomal DNA amplification for phylogenetic study *Journal of Bacteriology* **173**, 697-703.
Williams LA, Reimers C (1983) Role of bacterial mats in oxygen-deficient marine basins and coastal upwelling regimes: preliminary report. *Geology* **11**, 267-269.
Young KD (2006) The selective value of bacterial shape. *Microbiol. Mol. Biol. Rev.* **70,** 660–703.

7. Table Captions

Table 1: Primer sequences used for 16S rRNA PCR screening of MDA amplified product. Primers are specific to *Beggiatoa*, Bacterial, and Archaeal 16S rRNA amplification.

Name	Sequence (5'-3')	Target	Journal
233f	CCTATGCCGGATTAGCTTG	Beggiatoa 16S	Salman et al. 2011
ITSReub	GCCAAGGCATCCACC	Beggiatoa 16S	Salman et al. 2011
27f	AGAGTTTGATCMTGGCTCAG	Bacterial 16S	Weisburg et al. 1990
1492r	GGTTACCTTGTTACGACTT	Bacterial 16S	Weisburg et al. 1990
Arc8f	TCCGGTTGACCTGCC	Archaeal 16S	Sorensen & Teske 2007
9rts	CCCGCCAATTYCTTTAAGTTT	Archaeal 16S	Sorensen & Teske 2007

Table 2: Computing resources. “Labs” and Itasca servers used through Minnesota Supercomputing Institute (MSI), Symbiosys through the Bailey Geobiology Lab, and Geo-omics through the Michigan Geomicrobiology Lab

Cluster/Server Name	Processors	Total RAM	Per Node
Labs	808 24-core Intel Xeon 2.67 GHz	N/A	1 24-core Intel Xeon 2.67 GHz, 128 GB main memory
Itasca	2186 quad-core 2.8 GHz Intel Xeon X5560 “Nehalem EP”	24 TB main memory	2 quad-core 2.8 GHz Intel Xeon X5560 “Nehalem EP”, 24 GB main memory
Symbiosys	4 8-core 2.8 GHz Opteron 6320	512 GB main memory (32x16 GB)	N/A
Geo-omics	110 processors, 8 nodes	1142 GB main memory	16 processors, 23 GB main memory

Table 3: Comparison of Assemblies. MetaVelvet was utilized in subsequent analysis, due to lack of chimeric assemblies, quality of long (>10 kbp) contigs, and efficacy in prior genetic assays.

Assembly Method	Number of contigs	Longest contig	Number of contigs >1k bps	Number of contigs >10k bps	Mean contig size	N50 contig length
SGA	1299018	89168	11030 (0.8%)	772 (0.1%)	157	131
ABySS	1428195	217392	15405 (1.1%)	1418 (0.1%)	155	127
IDBA-UD	73666	153674	16928 (23.0%)	1787 (2.4%)	1379	5932
MetaVelvet + minimus2	144811	110107	12677 (>2500 bps) (8.7%)	1791 (1.2%)	1115 (6451 after 2500 bp limit)	2571

Table 4: Conserved Open Reading Frames (ORFs). Comparison of ORFs against IMG database used to determine phylogenetic affiliation of binned sequences.

Gene ID	Locus Tag	Gene Product Name	Genome
2528559721	Tmarg_01941	30S ribosomal protein S13	Thiomargarita nelsonii
2528557957	Tmarg_00175	Acetyltransferases, including N-acetylases of ribosomal proteins	Thiomargarita nelsonii
2528558427	Tmarg_00646	Acetyltransferases, including N-acetylases of ribosomal proteins	Thiomargarita nelsonii
2528559968	Tmarg_02188	Acetyltransferases, including N-acetylases of ribosomal proteins	Thiomargarita nelsonii
2528562931	Tmarg_05152	Acetyltransferases, including N-acetylases of ribosomal proteins	Thiomargarita nelsonii
2528558897	Tmarg_01117	LSU ribosomal protein L13P	Thiomargarita nelsonii
2528559724	Tmarg_01944	LSU ribosomal protein L15P	Thiomargarita nelsonii
2528558505	Tmarg_00724	LSU ribosomal protein L16P	Thiomargarita nelsonii
2528563708	Tmarg_05930	LSU ribosomal protein L17P	Thiomargarita nelsonii
2528559726	Tmarg_01946	LSU ribosomal protein L18P	Thiomargarita nelsonii
2528558780	Tmarg_00999	LSU ribosomal protein L20P	Thiomargarita nelsonii
2528558507	Tmarg_00726	LSU ribosomal protein L22P	Thiomargarita nelsonii
2528559463	Tmarg_01683	LSU ribosomal protein L23P	Thiomargarita nelsonii
2528558504	Tmarg_00723	LSU ribosomal protein L29P	Thiomargarita nelsonii
2528558509	Tmarg_00728	LSU ribosomal protein L2P	Thiomargarita nelsonii
2528560556	Tmarg_02776	LSU ribosomal protein L32P	Thiomargarita nelsonii
2528562905	Tmarg_05126	LSU ribosomal protein L33P	Thiomargarita nelsonii
2528559759	Tmarg_01979	LSU ribosomal protein L34P	Thiomargarita nelsonii
2528558781	Tmarg_01000	LSU ribosomal protein L35P	Thiomargarita nelsonii
2528557875	Tmarg_00093	LSU ribosomal protein L3P	Thiomargarita nelsonii
2528557876	Tmarg_00094	LSU ribosomal protein L4P	Thiomargarita nelsonii
2528559727	Tmarg_01947	LSU ribosomal protein L6P	Thiomargarita nelsonii
2528560588	Tmarg_02808	LSU ribosomal protein L9P	Thiomargarita nelsonii
2528559254	Tmarg_01474	Ribosomal protein L19	Thiomargarita nelsonii
2528560046	Tmarg_02266	ribosomal protein L21	Thiomargarita nelsonii
2528560045	Tmarg_02265	ribosomal protein L27	Thiomargarita nelsonii
2528559251	Tmarg_01471	ribosomal protein S16	Thiomargarita nelsonii
2528560397	Tmarg_02617	Ribosomal protein S20	Thiomargarita nelsonii
2528563367	Tmarg_05589	SSU ribosomal protein S10P	Thiomargarita nelsonii
2528559720	Tmarg_01940	SSU ribosomal protein S11P	Thiomargarita nelsonii
2528557871	Tmarg_00089	SSU ribosomal protein S12P	Thiomargarita nelsonii
2528558503	Tmarg_00722	SSU ribosomal protein S17P	Thiomargarita nelsonii
2528559240	Tmarg_01460	SSU ribosomal protein S18P	Thiomargarita nelsonii
2528560587	Tmarg_02807	SSU ribosomal protein S18P	Thiomargarita nelsonii
2528558508	Tmarg_00727	SSU ribosomal protein S19P	Thiomargarita nelsonii
2528558506	Tmarg_00725	SSU ribosomal protein S3P	Thiomargarita nelsonii
2528559719	Tmarg_01939	SSU ribosomal protein S4P	Thiomargarita nelsonii
2528559725	Tmarg_01945	SSU ribosomal protein S5P	Thiomargarita nelsonii
2528560586	Tmarg_02806	SSU ribosomal protein S6P	Thiomargarita nelsonii
2528557872	Tmarg_00090	SSU ribosomal protein S7P	Thiomargarita nelsonii
2528558898	Tmarg_01118	SSU ribosomal protein S9P	Thiomargarita nelsonii

Table 5: Putatively transferred proteins from cyanobacteria. Identified with best BLAST match to genes present in IMG database. First outlined section highlights genes encoding proteins relevant to transport of secondary metabolites, the second to genes relevant to cell wall biogenesis.

Gene ID	Locus Tag	Genome	COG
2528569331	Tmarg_11555	Thiomargarita nelsonii	COG1250===3-hydroxyacyl-CoA dehydrogenase
2528565971	Tmarg_08194	Thiomargarita nelsonii	COG2084===3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases
2528567130	Tmarg_09353	Thiomargarita nelsonii	COG3161===4-hydroxybenzoate synthetase (chorismate lyase)
2528565546	Tmarg_07769	Thiomargarita nelsonii	COG0156===7-keto-8-aminopelargonate synthetase and related enzymes
2528566485	Tmarg_08708	Thiomargarita nelsonii	COG1682===ABC-type polysaccharide/polyol phosphate export systems, permease component
2528569329	Tmarg_11553	Thiomargarita nelsonii	COG1960===Acyl-CoA dehydrogenases
2528568399	Tmarg_10622	Thiomargarita nelsonii	COG1051===ADP-ribose pyrophosphatase
2528564024	Tmarg_06247	Thiomargarita nelsonii	COG0367===Asparagine synthase (glutamine-hydrolyzing)
2528563938	Tmarg_06161	Thiomargarita nelsonii	COG1192===ATPases involved in chromosome partitioning
2528561111	Tmarg_03331	Thiomargarita nelsonii	COG3321===Polyketide synthase modules and related proteins
2528567564	Tmarg_09787	Thiomargarita nelsonii	COG1192===ATPases involved in chromosome partitioning
2528571007	Tmarg_13232	Thiomargarita nelsonii	COG1028===Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)
2528560792	Tmarg_03012	Thiomargarita nelsonii	COG0476===Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2
2528558877	Tmarg_01096	Thiomargarita nelsonii	COG1819===Glycosyl transferases, related to UDP-glucuronosyltransferase
2528564361	Tmarg_06584	Thiomargarita nelsonii	COG0438===Glycosyltransferase
2528558626	Tmarg_00845	Thiomargarita nelsonii	COG0463===Glycosyltransferases involved in cell wall biogenesis
2528570030	Tmarg_12255	Thiomargarita nelsonii	COG3882===Predicted enzyme involved in methoxymalonyl-ACP biosynthesis
2528570082	Tmarg_12307	Thiomargarita nelsonii	COG1091===dTDP-4-dehydrorhamnose reductase
2528565975	Tmarg_08198	Thiomargarita nelsonii	COG4948===L-alanine-DL-glutamate epimerase and related enzymes of enolase superfamily
2528571112	Tmarg_13337	Thiomargarita nelsonii	COG2227===2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase
2528570392	Tmarg_12617	Thiomargarita nelsonii	COG0451===Nucleoside-diphosphate-sugar epimerases
2528568277	Tmarg_10500	Thiomargarita nelsonii	COG1020===Non-ribosomal peptide synthetase modules and related proteins
2528557997	Tmarg_00216	Thiomargarita nelsonii	COG0451===Nucleoside-diphosphate-sugar epimerases
2528571051	Tmarg_13276	Thiomargarita nelsonii	COG0284===Orotidine-5'-phosphate decarboxylase
2528560041	Tmarg_02261	Thiomargarita nelsonii	COG0452===Phosphopantothenoylcysteine synthetase/decarboxylase
2528558627	Tmarg_00846	Thiomargarita nelsonii	COG1216===Predicted glycosyltransferases
2528567962	Tmarg_10185	Thiomargarita nelsonii	COG2161===Antitoxin of toxin-antitoxin stability system
2528564665	Tmarg_06888	Thiomargarita nelsonii	COG1280===Putative threonine efflux protein
2528558521	Tmarg_00740	Thiomargarita nelsonii	COG1087===UDP-glucose 4-epimerase

Table 6: Putatively transferred proteins from proteobacteria. Identified with best BLAST match to genes present in IMG database. The outlined sections highlight protein-encoding genes potentially transferred from Alpha-, Beta-, and Deltaproteobacteria, respectively.

Gene ID	Locus Tag	Genome	COG
2528567737	Tmarg_09960	Thiomargarita nelsonii	COG1898===dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes
2528559080	Tmarg_01300	Thiomargarita nelsonii	COG4591===ABC-type transport system, involved in lipoprotein release
2528558946	Tmarg_01166	Thiomargarita nelsonii	COG0739===Membrane proteins related to metalloendopeptidases
2528569871	Tmarg_12096	Thiomargarita nelsonii	COG1089===GDP-D-mannose dehydratase
2528565265	Tmarg_07488	Thiomargarita nelsonii	COG0707===UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase
2528558976	Tmarg_01196	Thiomargarita nelsonii	COG0481===Membrane GTPase LepA
2528559173	Tmarg_01393	Thiomargarita nelsonii	COG2089===Sialic acid synthase
2528559264	Tmarg_01484	Thiomargarita nelsonii	COG0463===Glycosyltransferases involved in cell wall biogenesis
2528558945	Tmarg_01165	Thiomargarita nelsonii	COG1664===Integral membrane protein CcmA involved in cell shape determination
2528570363	Tmarg_12588	Thiomargarita nelsonii	COG1089===GDP-D-mannose dehydratase
2528559204	Tmarg_01424	Thiomargarita nelsonii	COG0744===Membrane carboxypeptidase (penicillin-binding protein)
2528570483	Tmarg_12708	Thiomargarita nelsonii	COG0438===Glycosyltransferase
2528564031	Tmarg_06254	Thiomargarita nelsonii	COG0399===Predicted pyridoxal phosphate-dependent enzyme
2528569604	Tmarg_11829	Thiomargarita nelsonii	COG1922===Teichoic acid biosynthesis proteins
2528571132	Tmarg_13357	Thiomargarita nelsonii	COG0677===UDP-N-acetyl-D-mannosaminuronate dehydrogenase
2528569961	Tmarg_12186	Thiomargarita nelsonii	COG1086===Predicted nucleoside-diphosphate sugar epimerases
2528569102	Tmarg_11326	Thiomargarita nelsonii	COG4591===ABC-type transport system, involved in lipoprotein release
2528571147	Tmarg_13372	Thiomargarita nelsonii	COG0381===UDP-N-acetylglucosamine 2-epimerase
2528569425	Tmarg_11649	Thiomargarita nelsonii	COG0399===Predicted pyridoxal phosphate-dependent enzyme
2528559930	Tmarg_02150	Thiomargarita nelsonii	COG1292===Choline-glycine betaine transporter
2528570484	Tmarg_12709	Thiomargarita nelsonii	COG1004===Predicted UDP-glucose 6-dehydrogenase
2528562016	Tmarg_04236	Thiomargarita nelsonii	COG0463===Glycosyltransferases involved in cell wall biogenesis

Table 7: Heat map of Glycosyltransferase-encoding genes in Gammaproteobacteria. Large, vacuolated sulfur bacteria are shown to contain more cell wall biogenesis proteins than their closely-related, smaller counterparts. *Thiomargarita* is shown to have nearly double the biogenesis proteins of *Beggiatoa* sp. Orange Guaymas.

Function ID	Name	*T. nelsonii*	*Beggiatoa* sp. *Orange Guaymas*	*Beggiatoa* sp. *PS*	*T. crunogena*	*Vibrio harveyi*
COG0275	Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis	1	1	1	1	1
COG0399	Glycosyltransferases involved in cell wall biogenesis	12	6	7	3	1
COG0463	Glycosyltransferases, probably involved in cell wall biogenesis	37	19	12	2	3
COG2982	Uncharacterized protein involved in outer membrane biogenesis	1	1	1	1	1
COG3017	Outer membrane lipoprotein involved in outer membrane biogenesis	3	1	1	0	1

Table 8: Single-copy ribosomal proteins present in *Thiomargarita nelsonii*. Presence or absence of these single-copy genes is thought to indicate relative completeness of an assembled genome. The presence of all 41 indicates that the *Thiomargarita* genome can be thought to be relatively (>90%) complete.

Gene ID	Locus Tag	Gene Product Name	Genome
2528559721	Tmarg_01941	30S ribosomal protein S13	Thiomargarita nelsonii
2528557957	Tmarg_00175	Acetyltransferases, including N-acetylases of ribosomal proteins	Thiomargarita nelsonii
2528558427	Tmarg_00646	Acetyltransferases, including N-acetylases of ribosomal proteins	Thiomargarita nelsonii
2528559968	Tmarg_02188	Acetyltransferases, including N-acetylases of ribosomal proteins	Thiomargarita nelsonii
2528562931	Tmarg_05152	Acetyltransferases, including N-acetylases of ribosomal proteins	Thiomargarita nelsonii
2528558897	Tmarg_01117	LSU ribosomal protein L13P	Thiomargarita nelsonii
2528559724	Tmarg_01944	LSU ribosomal protein L15P	Thiomargarita nelsonii
2528558505	Tmarg_00724	LSU ribosomal protein L16P	Thiomargarita nelsonii
2528563708	Tmarg_05930	LSU ribosomal protein L17P	Thiomargarita nelsonii
2528559726	Tmarg_01946	LSU ribosomal protein L18P	Thiomargarita nelsonii
2528558780	Tmarg_00999	LSU ribosomal protein L20P	Thiomargarita nelsonii
2528558507	Tmarg_00726	LSU ribosomal protein L22P	Thiomargarita nelsonii
2528559463	Tmarg_01683	LSU ribosomal protein L23P	Thiomargarita nelsonii
2528558504	Tmarg_00723	LSU ribosomal protein L29P	Thiomargarita nelsonii
2528558509	Tmarg_00728	LSU ribosomal protein L2P	Thiomargarita nelsonii
2528560556	Tmarg_02776	LSU ribosomal protein L32P	Thiomargarita nelsonii
2528562905	Tmarg_05126	LSU ribosomal protein L33P	Thiomargarita nelsonii
2528559759	Tmarg_01979	LSU ribosomal protein L34P	Thiomargarita nelsonii
2528558781	Tmarg_01000	LSU ribosomal protein L35P	Thiomargarita nelsonii
2528557875	Tmarg_00093	LSU ribosomal protein L3P	Thiomargarita nelsonii
2528557876	Tmarg_00094	LSU ribosomal protein L4P	Thiomargarita nelsonii
2528559727	Tmarg_01947	LSU ribosomal protein L6P	Thiomargarita nelsonii
2528560588	Tmarg_02808	LSU ribosomal protein L9P	Thiomargarita nelsonii
2528559254	Tmarg_01474	Ribosomal protein L19	Thiomargarita nelsonii
2528560046	Tmarg_02266	ribosomal protein L21	Thiomargarita nelsonii
2528560045	Tmarg_02265	ribosomal protein L27	Thiomargarita nelsonii
2528559251	Tmarg_01471	ribosomal protein S16	Thiomargarita nelsonii
2528560397	Tmarg_02617	Ribosomal protein S20	Thiomargarita nelsonii
2528563367	Tmarg_05589	SSU ribosomal protein S10P	Thiomargarita nelsonii
2528559720	Tmarg_01940	SSU ribosomal protein S11P	Thiomargarita nelsonii
2528557871	Tmarg_00089	SSU ribosomal protein S12P	Thiomargarita nelsonii
2528558503	Tmarg_00722	SSU ribosomal protein S17P	Thiomargarita nelsonii
2528559240	Tmarg_01460	SSU ribosomal protein S18P	Thiomargarita nelsonii
2528560587	Tmarg_02807	SSU ribosomal protein S18P	Thiomargarita nelsonii
2528558508	Tmarg_00727	SSU ribosomal protein S19P	Thiomargarita nelsonii
2528558506	Tmarg_00725	SSU ribosomal protein S3P	Thiomargarita nelsonii
2528559719	Tmarg_01939	SSU ribosomal protein S4P	Thiomargarita nelsonii
2528559725	Tmarg_01945	SSU ribosomal protein S5P	Thiomargarita nelsonii
2528560586	Tmarg_02806	SSU ribosomal protein S6P	Thiomargarita nelsonii
2528557872	Tmarg_00090	SSU ribosomal protein S7P	Thiomargarita nelsonii
2528558898	Tmarg_01118	SSU ribosomal protein S9P	Thiomargarita nelsonii

Table 9: Duplicates of single-copy ribosomal proteins present in *Thiomargarita nelsonii*. Duplications of the single-copy proteins indicates artifacts from tetra-ESOM binning, stringent meta-assembly parameters, and/or single-nucleotide polymorphisms (SNPs) present in the genome copies residing in *Thiomargarita*.

Function ID	Name	Thio. nelsonii
COG0012	Predicted GTPase, probable translation factor	4
COG0016	Phenylalanyl-tRNA synthetase alpha subunit	3
COG0048	Ribosomal protein S12	4
COG0049	Ribosomal protein S7	4
COG0052	Ribosomal protein S2	2
COG0080	Ribosomal protein L11	1
COG0081	Ribosomal protein L1	1
COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	1
COG0087	Ribosomal protein L3	3
COG0088	Ribosomal protein L4	3
COG0090	Ribosomal protein L2	3
COG0091	Ribosomal protein L22	3
COG0092	Ribosomal protein S3	3
COG0093	Ribosomal protein L14	2
COG0094	Ribosomal protein L5	1
COG0096	Ribosomal protein S8	1
COG0097	Ribosomal protein L6P/L9E	2
COG0098	Ribosomal protein S5	2
COG0099	Ribosomal protein S13	2
COG0100	Ribosomal protein S11	2
COG0102	Ribosomal protein L13	3
COG0103	Ribosomal protein S9	3
COG0124	Histidyl-tRNA synthetase	3
COG0184	Ribosomal protein S15P/S13E	2
COG0185	Ribosomal protein S19	3
COG0186	Ribosomal protein S17	3
COG0197	Ribosomal protein L16/L10E	3
COG0200	Ribosomal protein L15	2
COG0201	Preprotein translocase subunit SecY	2
COG0256	Ribosomal protein L18	2
COG0495	Leucyl-tRNA synthetase	3
COG0522	Ribosomal protein S4 and related proteins	2
COG0525	Valyl-tRNA synthetase	4
COG0533	Metal-dependent proteases with possible chaperone activity	1
COG0541	Signal recognition particle GTPase	1

8. Figure Captions

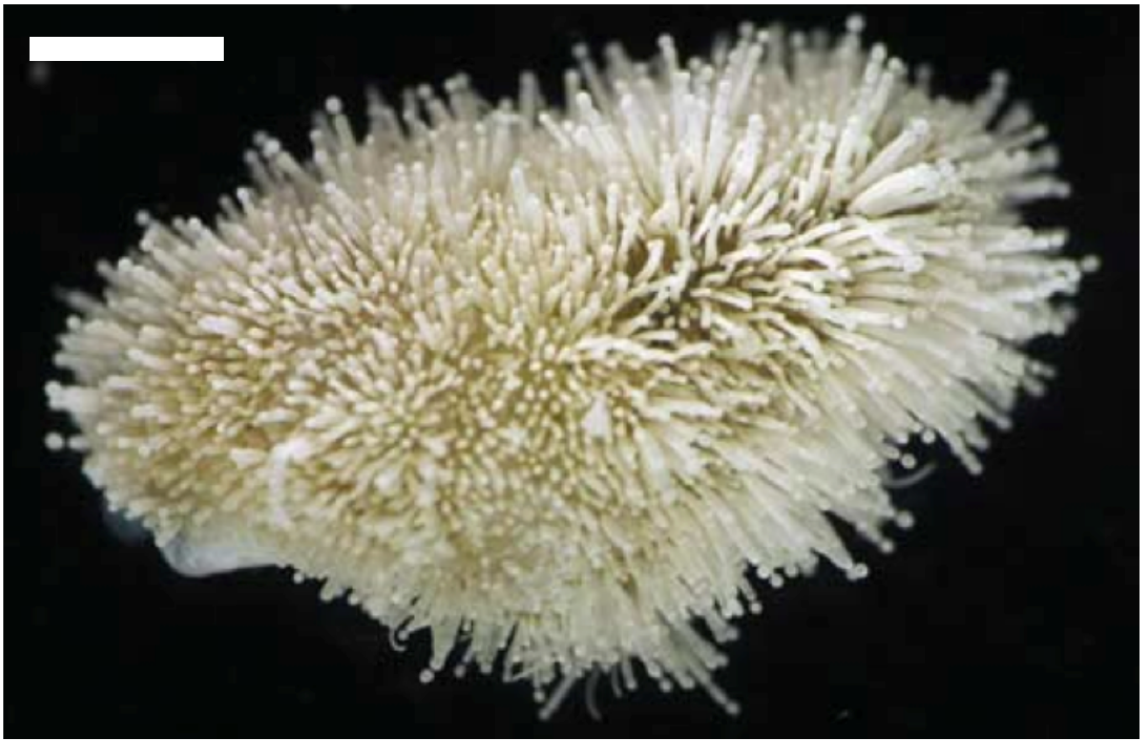

Figure 1: The microbial consortia present on *Provanna laevis*. Gastropods recovered from the Costa Rican margin and Hydrate Ridge. The elongate, immobile morphotypes of *Thiomargarita nelsonii*, *candidatus Maribeggiatoa* sp., and *Leucothrix* sp. form a fur-like coat. Scale bars represent 1 mm.

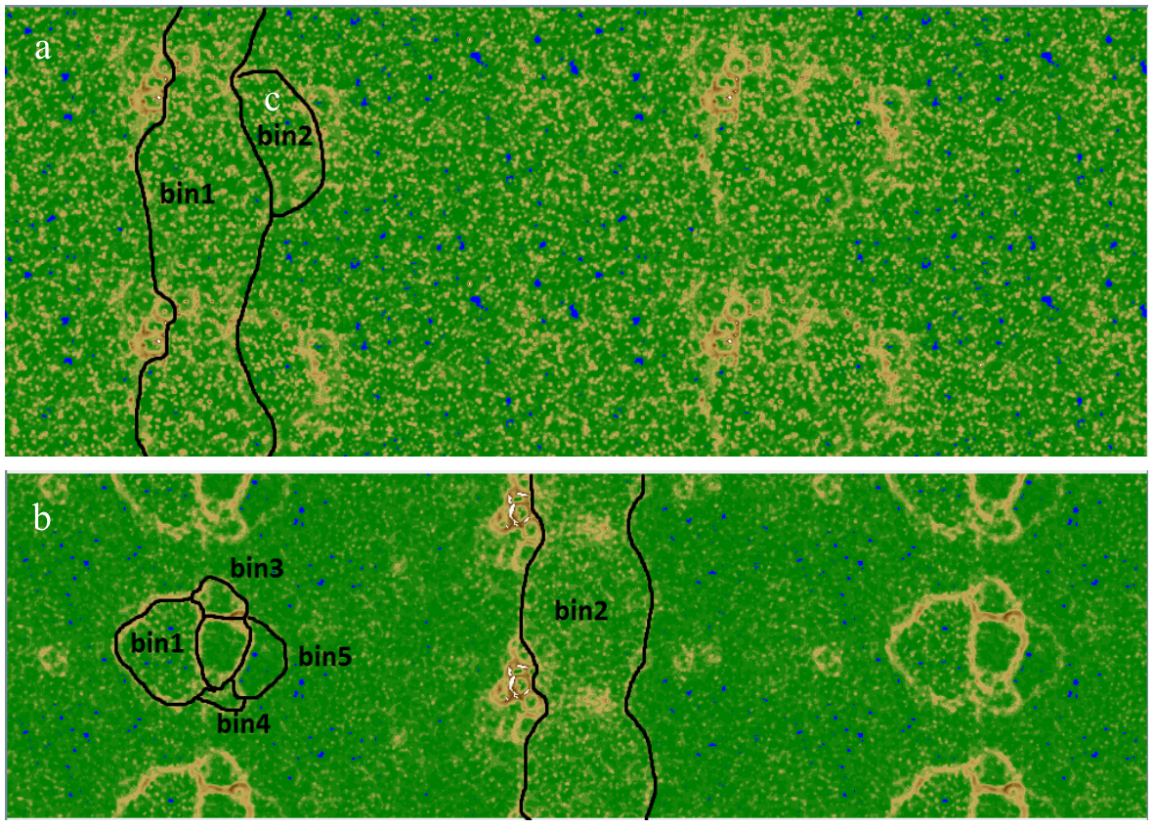

Figure 2: Emergent Self-Organizing Maps (ESOM) generated from sequences assembled with MetaVelvet. 2A shows the map generated from the sequences ranging from 2.5-5 kbps, 2B shows the map created with a range of 4-8 kbps. Bin 1 in Fig. 2A highlights the *Thiomargarita nelsonii* genome bin, with Bin 2 (2C) identified as a deltaproteobacterium.

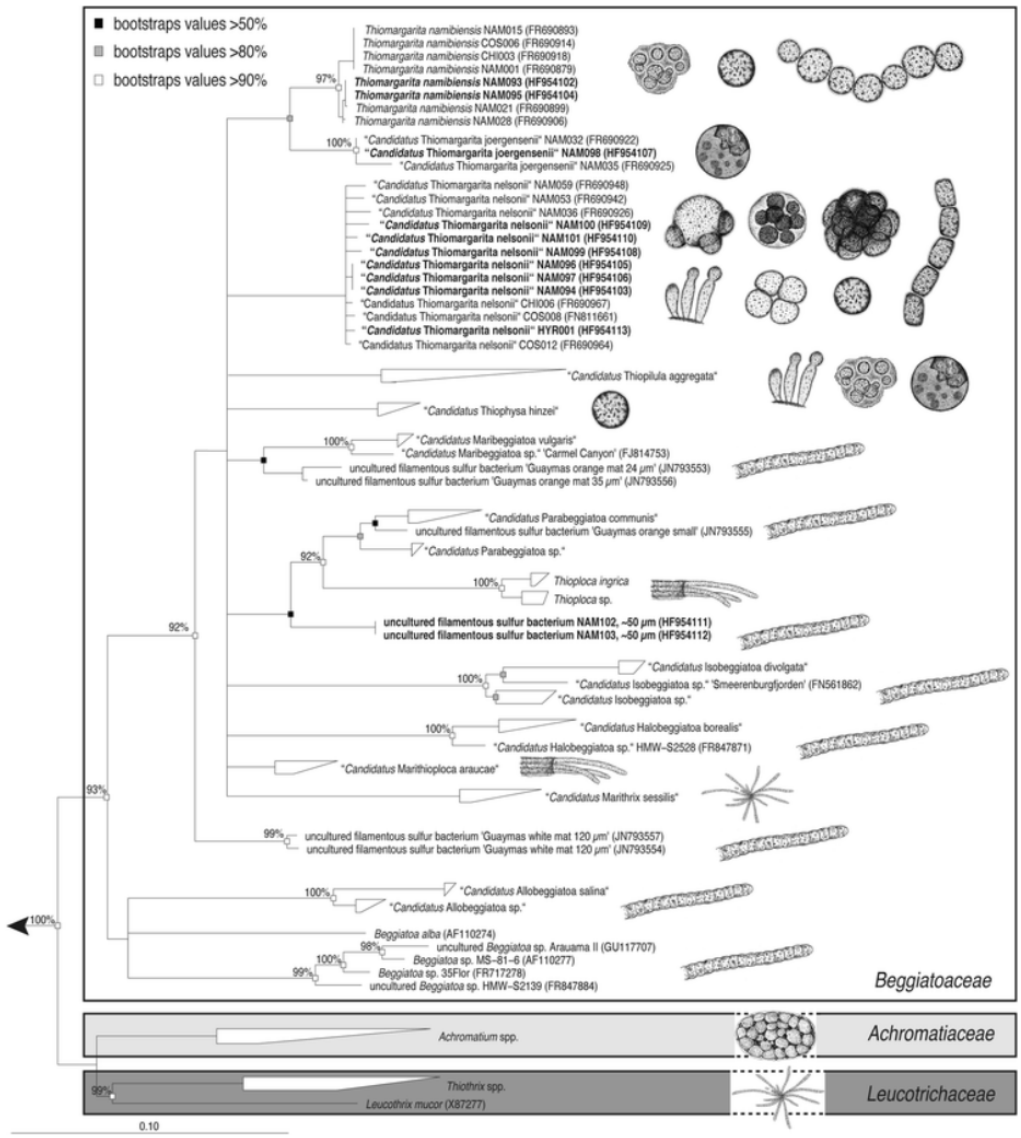

Figure 3: A Maximum-Likelihood tree of the 16S rRNA gene of the *Beggiatoaceae*, *Achromatiaceae*, and *Leucotrichaceae* in relation to the morphology of the more recently sequenced organisms in bold (Salman et al., 2013). Springer and *Antonie van Leeuwenhoek*, 104(2), 2013, pg. 178, Phylogenetic and morphologic complexity of giant sulphur bacteria, Salman, V., Bailey, J. V., & Teske, A, Figure 6, © Springer Science+Business Media Dordrecht 2013) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media. The final publication is available at: http://link.springer.com/article/10.1007%2Fs10482-013-9952-y

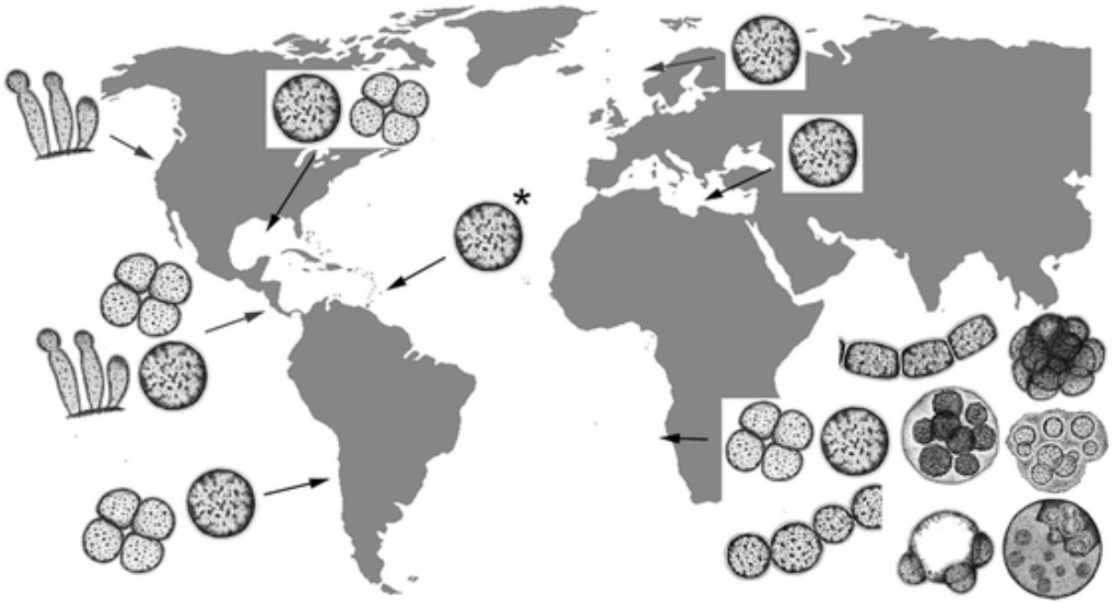

Figure 4: A comparative geographic morphotype analysis of currently-known non-filamentous sulfur bacteria. Springer and *Antonie van Leeuwenhoek*, 104(2), 2013, pg. 178, Phylogenetic and morphologic complexity of giant sulphur bacteria, Salman, V., Bailey, J. V., & Teske, A, Figure 5, © Springer Science+Business Media Dordrecht 2013) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media. The final publication is available at: http://link.springer.com/article/10.1007%2Fs10482-013-9952-y

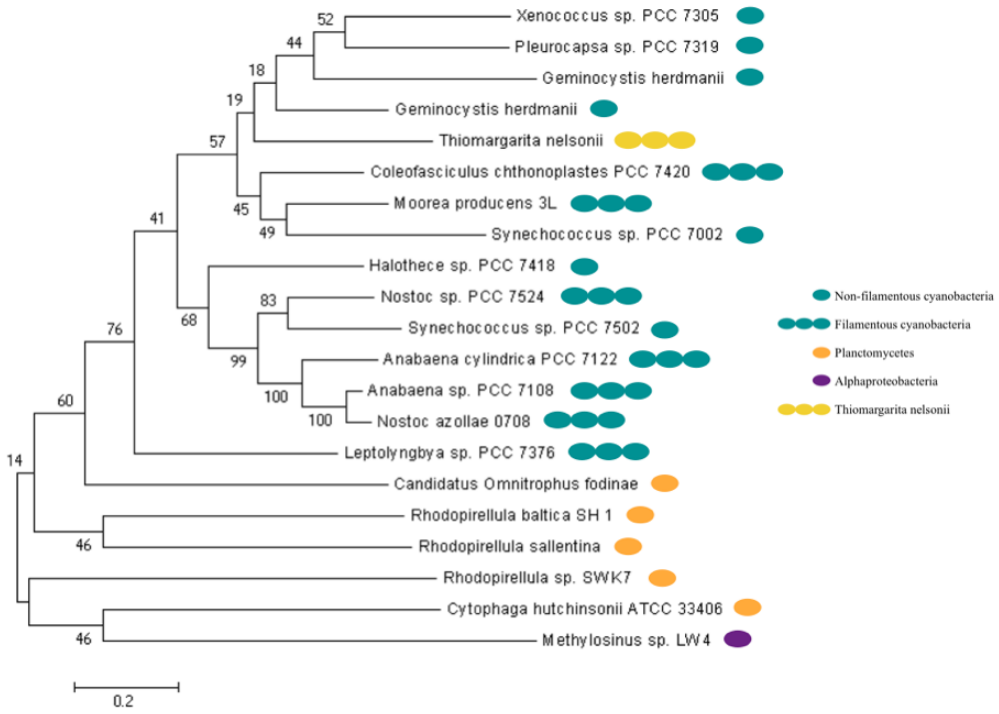

Figure 5: A maximum-likelihood phylogenetic tree of a putatively-transferred glycosyltransferase (Tmarg_03686) of the top 20 best protein BLAST (BLASTP) results. The tree shows a closest phylogenetic relationship to filamentous cyanobacteria and budding planctomycetes, providing a strong line of evidence for horizontal transfer of the gene in question.

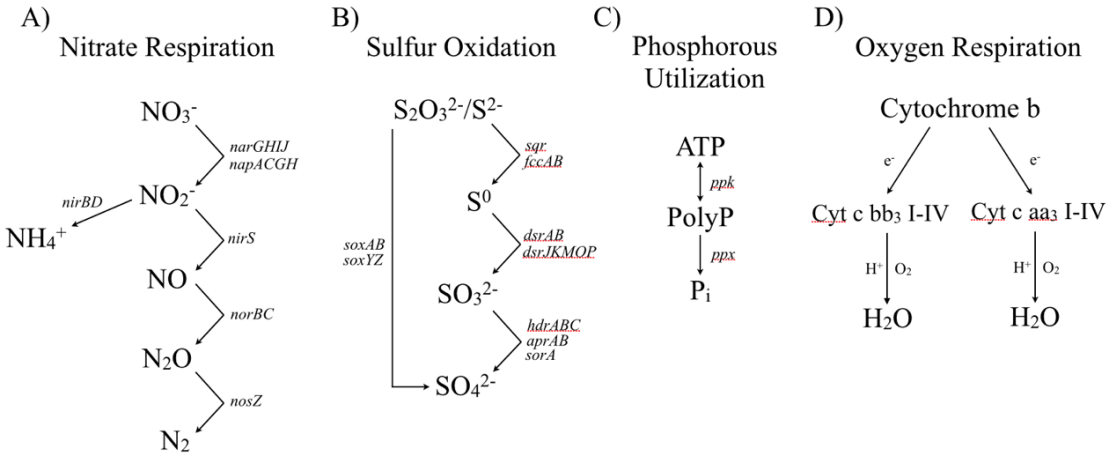

Figure 6: Predicted Nitrogen, Sulfur, Phosphorous, and Oxygen Metabolism of *Thiomargarita nelsonii*. Fig. 6A shows the full denitrification pathway and dissimilatory nitrate reduction to ammonia (DNRA) present. Enzymes reducing nitrite to ammonia and nitrous oxide to dinitrogen, the first shown in large, vacuolated sulfur bacteria, are shown here. Fig. 6B illustrates the encoded genes catalyzing oxidizing of sulfur species. Thiosulfate is likely oxidized via the Sox pathway. Fig. 6C shows the enzymes responsible for the creation and hydrolysis of polyphosphate. Fig. 6D shows the final steps in oxygen respiration, the depicted cytochrome c oxidases are observed to have differing oxygen affinities, with the bb_3 type having a higher affinity than the aa_3.

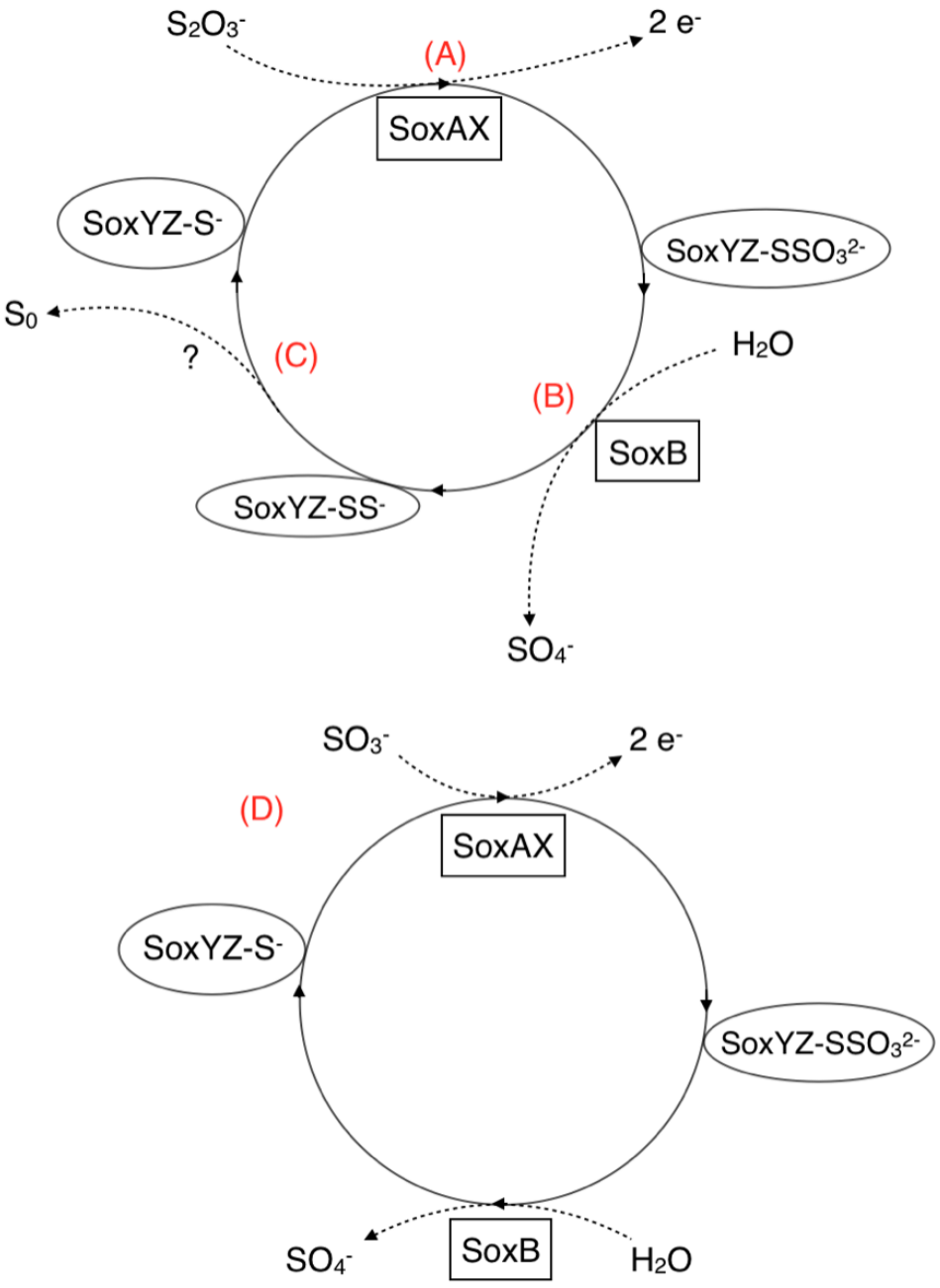

Figure 7: A schematic depicting the SOX system oxidizing thiosulfate (a-c) and sulfite (d) to sulfate. While the complete SOX system has been verified in *Paracoccus pantrophus* (Friedrich et al., 2001), the incomplete system shown here is capable of producing both sulfate and elemental sulfur observed in *Thiomargarita*.

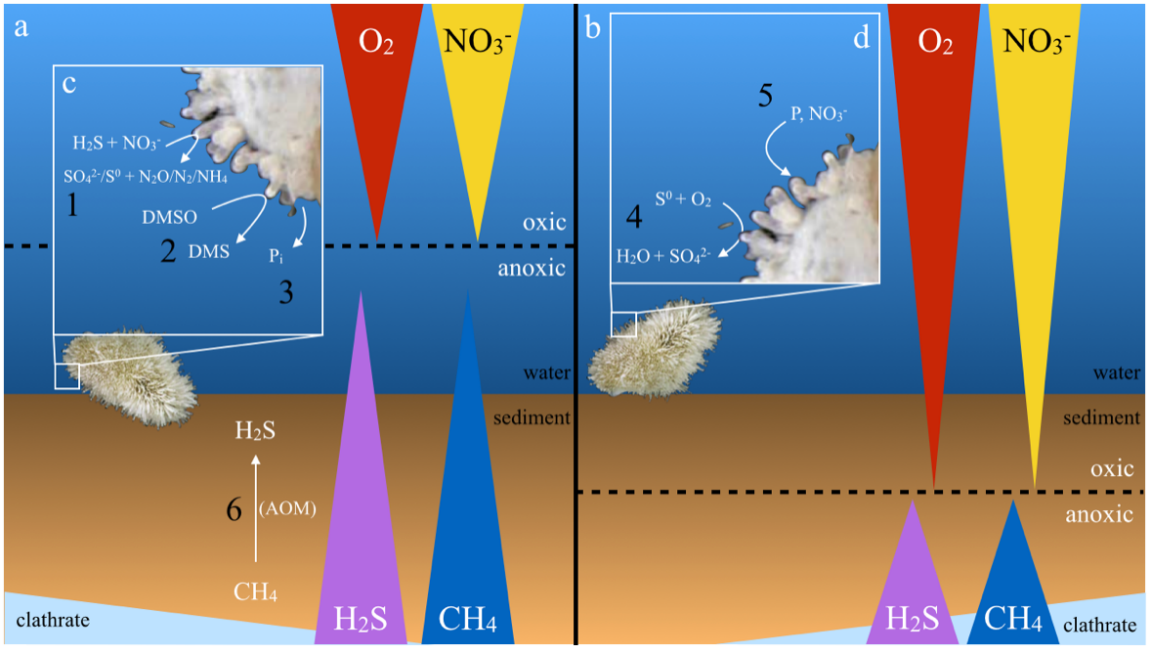

Figure 8: Scheme summarizing potential respiration pathways in host-attached *Thiomargarita* in varying vertical gradients of Oxygen, Nitrate, Sulfide, and Methane in a marine surface sediment. Fig. 8A illustrates the observed metazoan behavior and potential energy-yielding pathways in *Thiomargarita* when the oxic/anoxic boundary is above the sediment-water interface, and 8B illustrates the behavior and putative metabolic activity in oxygenated sediments.

1) Oxidation of Sulfide with intracellular Nitrate, producing Sulfate or Elemental Sulfur, and Nitrous Oxide, Dinitrogen, or Ammonia
2) Respiration of dimethyl sulfoxide (DMSO) to dimethyl sulfide (DMS) alternative to Nitrate and Oxygen
3) Hydrolysis of internally-stored polyphosphate granules to release phosphate into the local environment
4) Oxidation of membrane-bound sulfur globules to sulfate by Oxygen in sediments and water column
5) Uptake of phosphorous and nitrate from environment to store internally for times of nutrient stress (8A)