Elucidating the metabolic activities of SUP05, an abundant group of marine sulfur oxidizing gamma-proteobacteria

Katharine Marshall Lalish

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Robert M. Morris, Chair

John A. Baross

Anitra E. Ingalls

Program Authorized to Offer Degree:

School of Oceanography

University of Washington

Abstract

Elucidating the metabolic activities of SUP05, an abundant group of marine sulfur oxidizing gamma-proteobacteria

Katharine Marshall Lalish

Chair of the Supervisory Committee:
Professor Robert M. Morris
School of Oceanography

Chemoautotrophic bacteria that oxidize sulfur and reduce nitrogen in low oxygen zones carry out chemical transformations of molecules containing the key elements: carbon, nitrogen and sulfur. Members of the SUP05 clade of marine gamma-proteobacterial sulfur oxidizers (GSOs) are abundant and cosmopolitan in the oceans, consisting of metabolically diverse bacteria that have been found throughout the water column in oxic and anoxic marine environments and as symbionts in the gill tissues of deep sea clams and mussels. Environmental studies suggest that the SUP05 clade is comprised of diverse members with autotrophic, heterotrophic or mixotrophic capabilities. Although there are indications that these physiologies correspond to differing ecological niches, the lack of cultivated representatives has limited studies to elucidate patterns of metabolic diversity within the clade. In chapter one, I obtain the

first isolate from the SUP05 clade (*Thioglobus singularis* GSO-PS1) and show that energy from sulfur oxidation enhances heterotrophic biosynthesis in controlled growth experiments. In chapter two, I demonstrate that despite having a streamlined genome, *T. singularis* GSO-PS1 has the genetic potential for carbon mixotrophy and confirm that autotrophic and heterotrophic functions are expressed in the environment. In chapter three, I use the complete genome sequences of two SUP05 isolates, *T. singularis* GSO-PS1 and *T. autotrophica* SUP05-EF1 (Shah et al., in prep) to identify co-occurring subclades in diverse OMZs. Results from these analyses indicate that members of the *T. singularis* subclade are aerobic mixotrophs with the metabolic potential to produce energy and to biosynthesize using organic and inorganic substrates, and that members of the *T. autotrophica* subclade are chemoautotrophic facultative anaerobes with the metabolic potential for denitrification. This suggests that *T. singularis* and *T. autotrophica* represent SUP05 subclades with distinct ecological niches and with different roles in carbon, nitrogen, and sulfur cycling.

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGEMENTS

First I would like to thank my advisor, Robert Morris. It has been great doing science with you. Your optimism and encouraging words have gotten me through the tough parts of graduate school. I'm so fortunate to have gotten to be your first graduate student and it has been a pleasure seeing you learn and grow as a professor. You always make time to meet with me when I need guidance and your enthusiasm for data is marvelous. I've never seen anyone so happy about clean growth curves. You have been wonderful to work with and I wish you and your many graduate students to come all the best.

Thank you to my committee members, Jody Deming, John Baross, Anitra Ingalls, and Elhanan Borenstein. One of the things I struggle with is being articulate and confident despite being nervous about public speaking. I so appreciate our meetings every 6 months because it helped me fight this phobia. Your questions and suggestions were also a great help to my research.

Thank you to the members of the CEG. Especially Michael Carlson, my forever officemate and Vaughn Iverson, Jaci Saunders, and Gwenn Hennon, my bonus officemates. I cherish all of or office discussions, science and otherwise. To Vega Shah and Rachel Lange, my fellow lab mates: Vega, your culturing skills blow me away, Rachel, may GSO-PS1 be kinder to you than it was to me.

Thank you to my Seattle friends. Seattle has become home because of all of you. I especially want to thank Ellie Wytychak and Jen Manuzak, my Occidental College buddies. While I miss our dinners at the Marketplace, picnics by Lake Washington are even better.

Thank you to my parents John Marshall and Beth Marshall. I know that I always have your support and love and that has helped me so much over the years. Thank you to my sisters Lisa and Jackie Marshall. Everyone should have at least two sisters. We're like different versions of the same person and that is awesome.

Thank you to my daughter Lydia Lalish, your smiles make my heart flutter and give my stomach butterflies, I love you so much.

And last, thank you to my husband, Emmett Lalish. You've been so supportive since I met you 6 years ago. I am lucky to get to spend my life with you and I can't wait for our next adventure.

DEDICATION

To Lydia: be good, be kind, be strong.

INTRODUCTION

Marine microorganisms have critical roles in determining the chemical composition of seawater. Basic roles are often assigned to microorganisms to organize the mechanisms of biologically mediated nutrient cycles. For example, photoautotrophs are primary producers that use light for energy and heterotrophs are consumers of dissolved organic matter in seawater. Sulfur reducers use sulfate as their terminal electron acceptors reducing it to hydrogen sulfide and sulfur oxidizers make energy by oxidizing hydrogen sulfide to sulfate. These simple designations are warranted because they illustrate the major impact of these groups on nutrient pools and can be used to explain the biological contribution to major chemical inputs and outputs. However, these simplifications do not account for the significant metabolic diversity observed within many taxa. This dissertation defines the metabolic potential of the SUP05 clade of sulfur oxidizing gamma-proteobacteria. More specifically, it identifies notable metabolic diversity within the SUP05 clade and defines subclades with different functional roles in sulfidic and oxygen minimum zones (OMZs).

Sulfur is a critical element required for all life and an important energy source in sulfidic regions. The cycling of sulfur by microorganisms also impacts the carbon, oxygen and nitrogen cycles. The presence of oxygen oxidizes sulfur compounds to sulfate, thus making it the most abundant form of sulfur in today's ocean (1). Sulfur impacts oxygen through pyrite burial, which is formed in the absence of oxygen when microbes reduce sulfate to sulfide. The sulfide combines with iron and is buried, exiting the sulfur cycle, potentially for millions of years. This reaction is a source of oxygen to the marine environment because the sulfur is reduced before being buried, thus producing oxygen (2, 3). Sulfate reduction tends to be found coupled with

organic carbon respiration. Depending on the metabolic pathway of a microorganism, sulfur oxidation can be coupled with nitrogen or oxygen reduction and organic carbon respiration or primary production (4). Small shifts in marine microbial populations and changes to marine chemistry that occur over millions of years can strongly impact the chemical nature of the ocean and atmosphere.

Sulfur oxidizing bacteria have different metabolic pathways, giving them important, yet varying, roles in the carbon, nitrogen and sulfur cycles (4). Some fix carbon dioxide using energy from sulfur oxidation, others respire organic matter and use sulfur oxidation as a supplemental source of energy. Sulfur-oxidizing bacteria can also participate in complete or incomplete denitrification, producing nitrogen gas as an end product or providing reduced intermediates for other microorganisms (5). Sulfur-oxidizing organisms can also oxidize different sulfur compounds, creating intermediates with varying oxidation states.

Although many exceptions exist, the sulfur oxidation systems of various bacteria can be broken down into three basic groups (4). There are microbes capable of complete oxidation of sulfide to sulfate that have a full set of sulfur oxidation genes (sox) including a core set (soxABXYZ) and 11 supplementary genes (soxCDEFGHTRSVW). These are usually alpha-proteobacteria that fully oxidize sulfur and release eight electrons. They do no typically produce sulfur globule intermediates. Other microbes oxidize sulfur via formation of a tetrathionate intermediate and either have no sox genes or incomplete sox gene sets. These organisms tend to be beta and gamma-proteobacteria and they often produce sulfur globules. Some microbes use the branched thiosulfate oxidation pathway to oxidize reduced sulfur to elemental sulfur and deposit it as sulfur globules (internally or externally) via an incomplete sox set (*soxABXYZ*). These organisms lack *soxC* and *soxD*, which catalyze the oxidation of elemental sulfur and

prevent the production of elemental sulfur globules (6). A separate *dsr* system can further oxidize elemental sulfur globules. Organisms with this pathway include gamma- and beta-proteobacteria and bacteriodetes. Sulfur oxidation is an ancient metabolic strategy (4). Many different phyla are capable of sulfur oxidation, which makes it difficult to predict the potential of individual taxa.

My research focuses on the SUP05 clade, an abundant and cosmopolitan group of sulfur oxidizing marine bacteria. Members of the SUP05 clade were first identified in the tissues of clams and mussels living at hydrothermal vents and cold seeps (7, 8). 16S rDNA analyses indicated that these symbionts were closely related to thiotrophic bacteria that obtain energy from the oxidation of reduced sulfur (9). Like other symbiotic relationships in sulfide rich marine ecosystems, the host provides access to reduced sulfur sources for autotrophic symbionts and the symbionts provide the host with a source of organic carbon (9).

Abundant and free-living bacteria from the SUP05 clade that share sulfur oxidation genes with their symbiont relatives have since been identified in diverse seawater samples. Free-living members from this clade accounted for 37% of the bacterial 16S rDNA recovered from an oxygen minimum zone (OMZ) in a fjord in British Columbia and 45% of the 16S rDNA clones recovered from an OMZ in the South Atlantic (6, 10). Up to 90% of the bacteria in samples from the Suiyo Seamount caldera were from the SUP05 clade (11). They have also been found in the Yellow Sea, Monterey Bay, the Arabian Sea, the Benguela upwelling region, Suiyo Seamount, Axial Seamount, the Black Sea, the Arctic, and the Antarctic (10-18). These data suggest that the SUP05 clade is comprised of a diverse and abundant group of marine sulfur oxidizing bacteria that have adapted to both symbiotic and free-living lifestyles.

Previous phylogenetic analysis of 16S rDNA sequences suggested that the SUP05 clade

is composed of sister subclades (SUP05 and Arctic96BD-19) (6), though more recent analysis suggests that there may be more diversity within the group (Figure 1). Sequences related to SUP05 clone have been identified as symbiont and free-living bacteria that are dominant members of bacterial communities in OMZs. They have genes for sulfur oxidation, nitrogen reduction and autotrophy (6, 10, 17). Sequences related to Arctic96BD-19 are typically found in more oxygenated environments (>5 μM O_2) and appear to be ubiquitous autotrophic primary producers in oxygenated deep waters, potentially having a major impact on marine carbon cycling (18, 19). Arctic96BD-19 also have genes for sulfur oxidation, though less in known about their sulfur oxidation potential in oxygenated environments where most sulfur is fully oxidized.

Genomic data suggest that uncultured representatives from the SUP05 clades have the potential to oxidize reduced sulfur in oxic waters and waters lacking free oxygen (6, 10, 17-19). Members of the SUP05 subclade have genes that would allow them to oxidize hydrogen sulfide by reducing nitrogen species in the absence of free oxygen (6, 10, 17). In waters lacking free oxygen, hydrogen sulfide, in which sulfur has an oxidation state of -2, can accumulate to measurable concentrations (20). It is produced by sulfate reducing bacteria and by leaching from the ocean crust and is toxic to most forms of life (16). The sulfide is oxidized biotically by sulfur oxidizing bacteria and abiotically when anoxic sulfide rich waters are mixed with oxygenated surface waters. Though the extent of nitrogen reduction potential in SUP05 remains undefined, oxidation to nitrogen gas or nitrous oxide, would result in nitrogen removal from marine systems. Members of the SUP05 subclade were implicated in a major sulfide detoxification event associated with a highly sulfidic OMZ in the South Atlantic (10). Sulfur oxidation was previously thought to become important only in sulfidic OMZs, but Canfield et. al. found that

sulfur redox reactions were an important process in non-sulfidic OMZs and were strongly coupled to nitrogen redox reactions (21).

Members of Arctic96BD-19 subclade live in oxygenated waters, but contain sulfur oxidation genes. Little is known about biotic sulfur oxidation in oxygenated waters, where hydrogen sulfide is not present, but where less reduced sulfur may be available for further oxidation. Although sulfur is abundant in the oxic ocean, it is present as sulfate and fully oxidized with an oxidation state of +6. Other sulfur compounds with oxidation states of -2 to +6 are present at low levels throughout the ocean. In the 1970s, Tuttle and Jannasch identified bacteria capable of oxidizing thiosulfate in oxygenated waters and showed that thiosulfate can provide energy for autotrophy or heterotrophy. They concluded that widespread sulfur oxidation was an overlooked process (22, 23) because reduced sulfur is used rapidly.

Differences in the distribution of SUP05 and Arctic96BD-19 subclades suggests that they have different ecological niches that can be defined by differences in their metabolic capabilities and roles in the marine carbon, nitrogen and sulfur cycles. However, they share a high degree of sequence similarity (94% 16S rDNA sequence identity) and often co-occur. This has made it difficult to distinguish between the two and to assign functions based on environmental sequence data. This study reports the isolation of the first cultured representative from this lineage and uses complete genome sequences, physiology and environmental sequence data to elucidate their metabolic activities in diverse marine sulfidic and low oxygen areas. Results from this study indicate that the clade is comprised of chemoautotrophic (SUP05) and chemomixotrophic (Arctic96BD-19) subclades. The data support my overarching hypothesis that **carbon and energy acquisition determine the abundance and distribution of SUP05 subclades and their influence on marine nutrient cycles.**

Important SUP05 names

Though every effort was made to use clear names for the SUP05 clade, changes in our understanding of this clade have necessitated naming changes after the publication of chapter 1. Hopefully this guide will help if the naming becomes unclear.

GSO—gamma sulfur oxidizing proteobacteria, used to refer to the entire SUP05 clade in Ch. 1.
SUP05—the entire clade of bacteria, also used to refer to one of the sister subclades in Ch. 1.
SUP05/Arctic96BD-19—The original two SUP05 subclades.
T. singularis **GSO-PS1**—the type strain for the mixotrophic subclade, previously the Arctic96BD-19 subclade.
T. autotrophica **SUP05-EF1**—The type strain for the autotrophic subclade, previously the SUP05 subclade.

1. Zopfi J, Ferdelman TG, Fossing H (2004) Distribution and fate of sulfur intermediates—sulfite, tetrathionate, thiosulfate, and elemental sulfur—in marine sediments. *Geological Society of America Special Papers* 379:97-116.

2. Halevy I, Peters SE, Fischer WW (2012) Sulfate Burial Constraints on the Phanerozoic Sulfur Cycle. *Science* 337(6092):331–334.

3. Hurtgen MT (2012) The Marine Sulfur Cycle, Revisited. *Science* 337(6092):305–306.

4. Ghosh W, Dam B (2009) Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. *FEMS Microbiol Rev* 33(6):999–1043.

5. Lam P, Kuypers MMM (2011) Microbial Nitrogen Cycling Processes in Oxygen Minimum Zones. *Annu Rev Mar Sci* 3(1):317–345.

6. Walsh DA, et al. (2009) Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. *Science* 326(5952):578–582.

7. Distel DL, et al. (1988) Sulfur-oxidizing bacterial endosymbionts: analysis of phylogeny and specificity by 16S rRNA sequences. *J Bacteriol* 170(6):2506–2510.

8. Cavanaugh CM, Levering PR, Maki JS, Mitchell R, Lidstrom ME (1987) Symbiosis of methylotrophic bacteria and deep-sea mussels. *Nature* 325(6102):346–348.

9. Robinson JJ, Cavanaugh CM (1995) Expression of Form I and Form II Rubisco in Chemoautotrophic Symbioses: Implications for the Interpretation of Stable Carbon Isotope Values. *Limnol Oceanogr* 40(8):1496–1502.

10. Lavik G, et al. (2009) Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* 457(7229):581–584.

11. Sunamura M, Higashi Y, Miyako C, Ishibashi J, Maruyama A (2004) Two Bacteria Phylotypes Are Predominant in the Suiyo Seamount Hydrothermal Plume. *Appl Env Microbiol* 70(2):1190–1198.

12. Murray AE, Grzymski JJ (2007) Diversity and genomics of Antarctic marine micro-organisms. *Philos Trans R Soc Lond B Biol Sci* 362(1488):2259–2271.

13. Bano N, Hollibaugh JT (2002) Phylogenetic Composition of Bacterioplankton Assemblages from the Arctic Ocean. *Appl Environ Microbiol* 68(2):505–518.

14. Fuchs BM, Woebken D, Zubkov MV, Burkill P, Amann R (2005) Molecular identification of picoplankton populations in contrasting waters of the Arabian Sea. Available at: http://eprints.soton.ac.uk/16386/ [Accessed April 1, 2011].

15. Kato S, et al. (2009) Spatial distribution, diversity and composition of bacterial communities in sub-seafloor fluids at a deep-sea hydrothermal field of the Suiyo Seamount. *Deep Sea Res Part Oceanogr Res Pap* 56(10):1844–1855.

16. Anderson RE, Beltrán MT, Hallam SJ, Baross JA (2013) Microbial community structure across fluid gradients in the Juan de Fuca Ridge hydrothermal system. *FEMS Microbiol Ecol* 83(2):324–339. 15.

17. Zaikova E, et al. (2010) Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* 12(1):172–191.

18. Swan BK, et al. (2011) Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* 333(6047):1296 –1300.

19. Swan BK, et al. (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci*. Available at: http://www.pnas.org/content/early/2013/06/25/1304246110 [Accessed July 18, 2013].

20. Sievert SM, Kiene RP, Schultz-Vogt HN The sulfur cycle. Available at: https://darchive.mblwhoilibrary.org/handle/1912/2786 [Accessed April 14, 2011].

21. Canfield DE, et al. (2010) A Cryptic Sulfur Cycle in Oxygen-Minimum–Zone Waters off the Chilean Coast. *Science* 330(6009):1375 –1378.

22. Tuttle JH, Jannasch HW (1976) Microbial Utilization of Thiosulfate in the Deep Sea. *Limnol Oceanogr* 21(5):697–701.

23. Tuttle JH, Jannasch HW (1977) Thiosulfate stimulation of microbial dark assimilation of carbon dioxide in shallow marine waters. *Microb Ecol* 4(1):9–25.

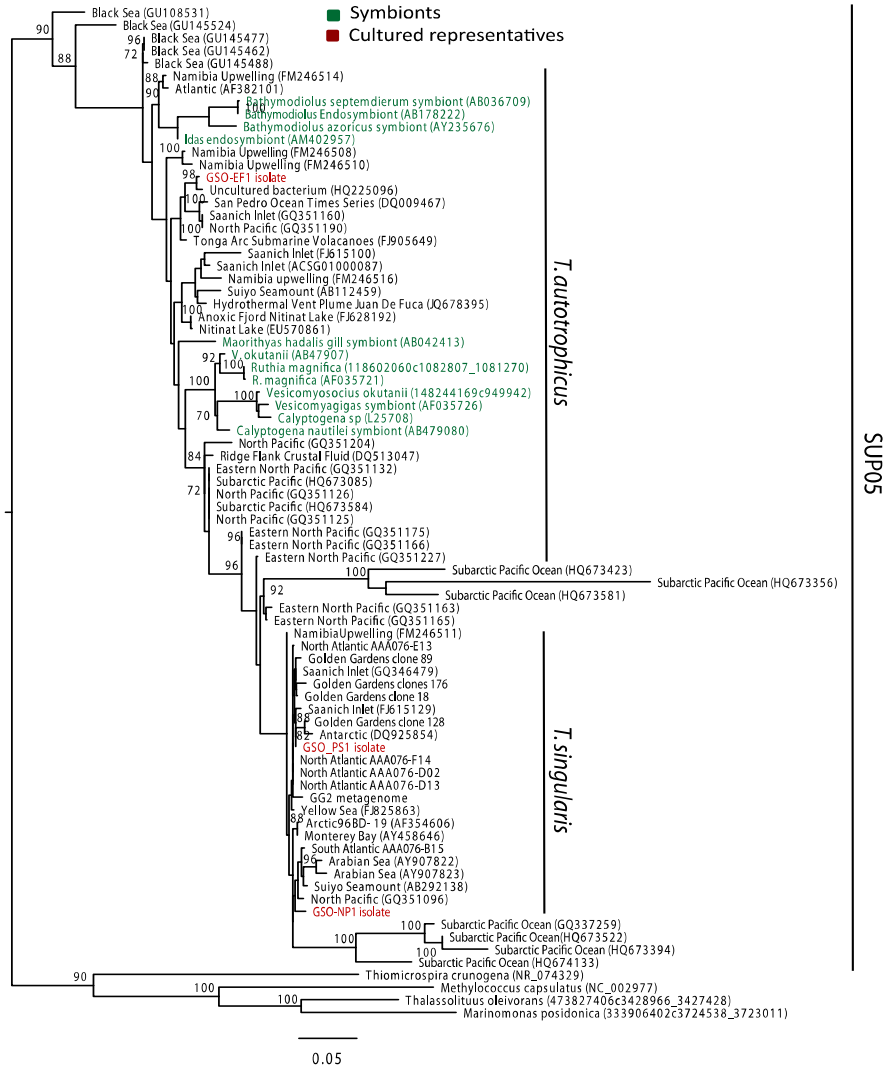

Figure 0.1. Phylogenetic analysis of SUP05 16S rRNA constructed using a maximum likelihood reference tree that contained nearly complete (1300 bp) sequences from cultures, environment clones and symbionts. Bootstrap values ≥70 (100 replicates) are reported above the nodes. Four outgroups were selected. SUP05/*T. singularis* clade designations (black bars) are based on the study by Walsh et al. (2009) [1].

Chapter 1. ISOLATION OF AN AEROBIC SULFUR OXIDIZER FROM THE SUP05/ARCTIC96BD-19 CLADE

Katharine T. Marshall and Robert M. Morris

ISME Journal (2013) 7(2):452–455.

Key words: Cultivation/GSO/Arctic96BD-19/SUP05/Sulfur

Category: Microbial ecology and functional diversity of natural habitats

1.1 ABSTRACT

Bacteria from the uncultured SUP05/Arctic96BD-19 clade of gamma proteobacterial sulfur oxidizers (GSO) have the genetic potential to oxidize reduced sulfur and fix carbon in the tissues of clams and mussels, in oxygen minimum zones, and throughout the deep ocean (>200 m). Here we report isolation of the first cultured representative from this GSO clade. Closely related cultures were obtained from surface waters in Puget Sound and from the deep chlorophyll maximum in the North Pacific gyre. Pure cultures grow aerobically on natural seawater media, oxidize sulfur, and reach higher final cell densities when glucose and thiosulfate are added to the media. This suggests that aerobic sulfur oxidation enhances organic carbon utilization in the oceans. The first isolate from the SUP05/Arctic96BD-19 clade was given the provisional taxonomic assignment “Candidatus: Thioglobus singularis”, alluding to the clade’s known role in sulfur oxidation and the isolate’s planktonic lifestyle.

1.2 INTRODUCTION

Marine sulfur oxidizing bacteria thrive in low oxygen environments. They have critical roles in the production of organic matter using energy derived from hydrogen sulfide that is produced by sulfate reducing bacteria or leached from the ocean crust at high temperatures (Sievert *et al*., 2007). The reduced sulfur in hydrogen sulfide has an oxidation state of -2 and is available as an energy source for anaerobic bacteria that oxidize sulfide by reducing nitrate to nitrogen gas, nitrous oxide, or ammonia. Less is known about sulfur oxidation in oxygenated marine waters. There were hints in the 1970's that this was an important process. Tuttle and Jannasch (1976; 1977) identified bacteria capable of oxidizing thiosulfate in diverse inshore and offshore waters and demonstrated the potential for thiosulfate oxidation to enhance carbon fixation and glucose utilization. More recently, Swan *et al.,* (2011) identified genes for sulfur oxidation, carbon fixation, and organic matter transport in the SUP05/Arctic96BD-19 clade of gamma proteobacterial sulfur oxidizers (GSO) from oxygenated waters in the dark ocean.

The SUP05/Arctic96BD-19 clade of marine GSOs is comprised of symbiotic and planktonic representatives from the "Basal" group of gamma proteobacteria, which contains the most ancient and highly diverged lineages of gamma proteobacteria (Williams *et al.*, 2010). Here we report isolation and thiosulfate enhanced heterotrophic growth of the first pure culture representative from the SUP05/Arctic96BD-19 clade of GSOs (Supplementary Methods). Several planktonic strains were cultured from surface waters in Puget Sound in November 2009 (n=3) and from the deep chlorophyll maximum (DCM) in the North Pacific gyre in 2011 (n=5) (Supplemental Table 1). The presence of viable cells at diverse sites in the North Pacific suggests that active members of this clade are widely distributed in the ocean surface layer. A representative strain was revived and purity was verified by observing a single 16S rRNA gene

fragment in successive cultures and imaging cells with an isolate-specific fluorescence *in situ* hybridization (FISH) probe (Supplementary Figure 1).

1.3 Results and Discussion

Cultured SUP05/Arctic96BD-19 cells are extremely small and produce extracellular globules (Figure 1). The potential for sulfur oxidation was confirmed by identifying sulfur S° in the extracellular globules using scanning transmission electron microscopy (Figure 1) combined with energy-dispersive x-ray spectroscopy (Figure 1B and 1C). These data support the genomic findings of Walsh *et al.*, (2009), which suggested that members of this clade have the capacity to oxidize and store S° due to an incomplete Sox pathway. Suspected roles in marine carbon and sulfur cycles were further evaluated using previously published PCR primers or new primers targeting genes involved in carbon fixation and sulfur oxidation (Swan *et al*., 2011; Blazejak *et al*., 2006). Of the primer pairs tried, only the dissimilatory *apr*A gene was identified (Blazejak *et al*., 2006; Hipp *et al*., 1997; Friedrich *et al*., 2005; Kelly *et al*., 1997). The inability to amplify carbon fixation genes using degenerate primers suggests that cultured GSOs are significantly diverged from known GSOs that occupy anoxic zones and the deep ocean (Walsh *et al.*, 2009; Swan *et al.*, 2011) or that they get all of their carbon from organic compounds.

Sulfur oxidizing bacteria capable of heterotrophy are well known. They are found in diverse lineages, including alpha and gamma proteobacteria and the *Cytophaga-Flavobacterium-Bacteroides* (CFB) group (Teske *et al.,* 2000) and some marine *Firmicutes*. The gamma proteobacteria include *Pseudomonads*, *Pseudoalteromonas* and *Halomonas-Deleya* (Tuttle *et al.,* 1974; Teske *et al.,* 2000). These lineages (along with epsilon proteobacteria) are common at hydrothermal vents and at oxic/anoxic interfaces, where reduced sulfur compounds are used to

produce energy (Podgorsek *et al*., 2004; Ruby *et al.,* 1981; Durand *et al.,* 1994; Tuttle *et al*., 1974; Teske *et al.,* 2000; Sorokin, 2003). Sulfur oxidation enables these bacteria to use more organic carbon for biosynthesis and less for respiration, giving them the ability to compete in a wide range of marine habitats (Podgorsek *et al.,* 2004; Teske *et al.,* 2000).

The marine sulfate reservoir has increased over the last 0.5-2.5 billion years, to its current concentration of 28 mM (Canfield *et al.*, 2000). Relative to sulfate, there is little to no accumulation of the inorganic sulfur cycle intermediates thiosulfate and tetrathionate, which suggests that ubiquitous sulfur oxidizing marine bacteria oxidize these compounds in seawater. We evaluated thiosulfate as a potential energy source for heterotrophic growth on glucose (Supplementary Methods). Thiosulfate is ubiquitous at low concentrations in seawater (<5 μM) and is relatively stable in the presence of oxygen (Hayes *et al*., 2006). Thiosulfate enhanced heterotrophic glucose utilization (Figure 2). Growth rates and final cell densities increased when glucose alone was added to the media and were highest when both glucose and thiosulfate were added to the media (Figure 2A and 2B, respectively). It is unlikely that this is due to co-limitation of carbon and sulfur because increasing concentrations of glucose alone increased cell densities, because cell densities at each concentration of glucose increased when thiosulfate was also added to the media, and because sulfur globules are only produced by sulfur oxidizing organisms. Combined, these data suggest that energy derived from sulfur oxidation enhances organic carbon turnover in the oxygenated ocean.

1.4 CONCLUSION

We propose the provisional taxonomic assignment "*Candidatus*: Thioglobus singularis", alluding the clade's known role in sulfur oxidation and the isolates planktonic lifestyle.

Thioglobus gen. nov.

Thioglobus singularis sp. nov.

Etymology. *Thio* (Gr. Noun): sulfur*, globus* (L. masc. noun): ball, sphere, globe. *Singularis* (L. adj.): alone, singular. The Genus name alludes to the clade's ability to oxidize sulfur and to the sulfur globules found on the outside of the cells. The species name alludes to the fact that this is a free-living member of the clade, rather than a symbiont.

Locality: surface waters in Puget Sound.

Diagnosis: a small mesophilic sulfur oxidizer from the gamma proteobacteria.

Accession numbers deposited in public databases

Gene sequences were assigned the following NCBI accession numbers; Puget Sound cultures, 16S rRNA genes (JQ254014–JQ254058); North Pacific gyre cultures, 16S rRNA genes (JQ253970–JQ254013); SUP05/Arctic96BD-19 isolate, Puget Sound isolate 16S rRNA gene (JN003574); SUP05/Arctic96BD-19 isolate, North Pacific gyre isolate 16S rRNA gene (JQ253969) and Puget Sound isolate *apr*A gene (JQ253968). Cultures will be provided upon request.

1.5 ACKNOWLEDGEMENTS

We thank the captain, officers and crew of the *RV Thompson* for their assistance during 2009 and 2011 cruises. This work was supported by the University of Washington, School of Oceanography, a grant from the National Science Foundation (OCE-0723866) and the Pacific Northwest Center for Human Health and Ocean Sciences (National Institute of Environmental Health: P50 ES012762 and National Science Foundation: OCE-0434087). We would also like to thank V. Iverson for help implementing a culture classification strategy.

1.6 ADDITIONAL INFORMATION

Additional information that was not included in this publication can be found in Appendix A including seasonal analysis of the relative abundance of GSOs, other cultures that were obtained

from the experiment that resulted in the GSO-PS1 culture, and the growth response of GSO-PS1 to a *Thalassiosira pseudonana* lysate amendment.

1.7 SUPPLEMENTARY METHODS

Bacterial cultures from Puget Sound and the North Pacific gyre

Members of the SUP05/Arctic96BD-19 clade of GSOs were cultured using a high-throughput dilution to extinction culturing approach modified from Connon and Giovannoni (2002). Culturing experiments were conducted with surface water collected from the Puget Sound main basin (47° 41.24' N, 122° 24.14' W) in November 2009 and from the DCM (45 m) in the North Pacific gyre near Axial seamount in August 2011. Culture media was prepared by pre-filtering seawater through a 0.8 μm polyethersulfone filter (Supor-200, Pall Corp, Ann Arbor, MI) and by sterilizing the filtrate using a 30 kD biomax polyethersulfone tangential flow filtration (TFF) cartridge (Millipore, Billerica, MA). TFF filter-sterilized seawater media was collected in autoclaved polycarbonate bottles and stored at *in situ* temperatures. Matching whole water samples were diluted (3-5 cells ml^{-1}) in TFF filter-sterilized seawater media and added to each well of an acid washed (10% HCL) 96 well Teflon plate (Sonomatesting, Forestville, CA).

Each experiment consisted of 576 cultures divided into two treatments. One treatment contained filter sterilized seawater media (unamended) and one contained seawater media amended with a natural source of organic carbon (lysate). Vitamins B1, B6, B7, and B12 were added to the North Pacific gyre lysate treatment at a final concentration of 10 nM each. Plates were incubated in the dark at *in situ* temperatures (Puget Sound, 13 °C and North Pacific, 10 °C) and screened for growth on an Easyflow Guava flowcytometer equipped with a 96 well plate reader (Millipore, Billerica, MA). Cultures were checked for growth by transferring 150 μL of culture to a new plate and by staining the cells with Syber Green I (Invitrogen, Carlsbad, CA)

diluted in TRIS buffer and at a final concentration of 1/2000, as previously described (Stingl *et al*., 2007).

Taxonomic assignments were determined for bacterial cultures that were positive for growth by extracting and amplifying the 16S rRNA gene. DNA from 200 µl of culture was extracted using a DNeasy Blood and Tissue Kit (QIAGEN, Germantown, MD, USA). 16S rRNA genes were amplified using a semi-nested PCR reaction with *Taq* polymerase (Fermentas, Hannover, MD, USA) and bacterial primers. Amplifications were performed in a C1000 thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA) using the following conditions: 35 cycles with 8F and 1492R primers followed by 38 cycles with 8F and 519R primers. The same conditions were used for each PCR reaction; denaturation at 94 °C for 30 s., annealing at 55 °C for one min, elongation at 72 °C for two min, and a final elongation step at 72°C for 10 min Amplicons were sequenced at the High-Throughput Genomics Unit (University of Washington, Seattle, WA, USA). Taxonomic assignments were determined using the Bayesian method of Wang *et al*., (2007) and a database augmented with sequences from marine environmental clades as previously described (Iverson *et al.*, 2012) (Supplementary Table 1).

Purity of SUP05/Arctic96BD-19 cultures

SUP05/Arctic96BD-19 culture purity was confirmed by T-RFLP analyses. Briefly, 16S rRNA genes were amplified using an 8F primer that was 5' end-labeled with the phosphoramidite fluorochrome 5-carboxy-fluorescein (6-FAM) and the following conditions: 32 cycles, annealing at 55 °C for 1 min, elongation at 72 °C for 2 min, and denaturation at 94°C for 30 s. Amplicons were restricted overnight at 37 °C with the enzyme *Bsu*RI (*Hae*III) and a 324 bp terminal

restriction fragment matching that predicted for the 16S rRNA gene sequence was resolved on an ABI 3730 DNA Analyzer (Supplementary Figure 1A).

SUP05/Arctic96BD-19 culture purity was further confirmed by fluorescence in situ hybridization (FISH). A Cy3 probe (GSO-1448R 5'-GGTGACCGTCCTCAATAAAG-3') was designed that exactly matched the isolate using a custom Arb database (Ludwig, *et al*., 2004). Probe specificity was checked using RDP probe match. Of 2.1 x 10^6 bacterial sequences in the database, 64 sequences matched the probe with 59 matching unclassified gamma proteobacteria belonging to the SUP05/Arctic96BD-19 clade. The five additional sequences include single hits to different lineages in the *Flavobacteriaceae* and alpha and gamma proteobacteria. Exponentially growing cultures were fixed overnight (9 ml) in 1% formaldehyde and filtered onto a 0.2 μm polycarbonate filter (Fisher, Pittsburgh, PA). Stringency was empirically determined by increasing the hybridization wash temperature. Hybridization reactions were performed on membrane sections at 37°C for 16 hours in hybridization buffer [900mM NaCl, 20mM Tris (pH 7. 4), 0.01% (w/v) sodium dodecyl sulphate (SDS), and 15% formamide, and the GSO-1448R Cy3-labeled oligonucleotide probe]. Control hybridization reactions were performed with a low stringency buffer containing 15% formamide and a Cy3-labeled negative control probe (338F). Probes had a final concentration of 2ng μl^{-1} for each hybridization reaction. Optimal hybridization stringency was achieved by hybridization washes [20mM Tris (pH 7. 4), 5 mM ethylenediaminetetraacetic acid (EDTA), 0. 01% SDS, and 150 mM NaCl] for two 10 min intervals. Nucleic acid staining occurred by mounting the filters in Citifluor (Ted Pella, Redding, CA) that contained 4', 6-diamidino-2-phenylindole (DAPI) at a final concentration of 5μg ml^{-1}. Slides were sealed and stored at -20°C.

Cy3-positive (FISH) and DAPI-positive cells were viewed using a Nikon 80i epifluorescence microscope equipped with a Photometrics CoolSNAPHQ2 digital camera, filter sets appropriate for Cy3 and DAPI, and the NIS-Elements Basic Research Acquisition and Analyses package (Nikon Instruments, Inc. Melville, NY). All DAPI images were segmented and overlain onto corresponding Cy3 image segmentations in order to identify positive probe signals with corresponding DAPI signals (Supplementary Figure 1B). All negative control images had no signal with the Cy3 filter set.

Gene sequences from SUP05/Arctic96BD-19 cultures

Identity was confirmed in a large volume transfer culture (1L) by filtering cells onto a sterile Supor-200 0.2 µm polyethersulfone filter (Pall, Port Washington, NY). Briefly, the filter was placed in a 2 ml cryovial, treated with 1 ml of lysis buffer, flash frozen in liquid nitrogen, and then stored at -20 ºC. The sample was later thawed and genomic DNA was extracted from the cell lysate using a DNeasy Blood and Tissue kit (QIAGEN, Germantown, MD). The 16S rRNA gene was amplified using bacterial primers, 8F and 1492R. Eight 38 cycle PCR reactions were pooled and purified, and the 16S rRNA gene was sequenced using primers 519F and R, 926F and R.

The sulfur oxidation gene *apr*A was sequenced by PCR amplification using aps1F (5'-TGGCAGATCATGATYMAYGG-3') and aps4R (5'-GCGCCAACYGGRCCRTA-3') (Blazejak *et al*., 2006). PCR reactions were performed using 38 cycles and published PCR conditions (Blazejak *et al*., 2006). Amplified DNA was cleaned using a Minelute kit (QIAGEN, Germantown, MD). All sequencing was done at the High-Throughput Genomics Unit (University of Washington, Seattle, WA).

Images of SUP05/Arctic96BD-19 cultures and energy dispersive x-ray spectroscopy

For scanning transmission electron microscopy (STEM), 1 L of exponentially growing cells were filter onto a 0.2 µm polycarbonate filter (Fisher, Pittsburgh, PA). 0.1 M sodium cacodylate with 2% EM grade gluteraldehyde was added to the top of the filter and left for 1 hour before being pulled through the filter with a vacuum pump. Increasing ethanol concentrations (20%, 35%, 50%, 75%, 95%) were added to the top of the filter and left for 10 minutes each before being pulled through the filter. Cells were washed off of the filter into a microfuge tube with 1 ml of 100% ethanol and spun for 1 hour at 14 000 rpm on a Sorvall RC-5B (Du Pont Instruments, Belle, WV). The supernatent was removed and the cell pellet was resuspended in 10 µl of 80% ethanol. 5 µl of sample was placed on a copper grid and dried for 24 hours before visualization and composition analysis at the Nanotech User Facility (University of Washington, Seattle, WA) using a Tecnai G2 F20 microscope equipped with EDX (FEI, Hillsboro, OR).

Growth of SUP05/Arctic96BD-19 cultures on organic carbon and thiosulfate

Pure culture growth experiments were conducted in 30 kD filter sterilized seawater media collected from 10 m in the North Pacific gyre (August 2010). Treatments were conducted in 6 replicated wells, with 1.5 ml per well in an acid washed 96 well Teflon plate. Treatments included seawater media, and media amended with either diatom lysate, 1 mM thiosulfate, 1, 10 and 100 µM of glucose, or 1, 10, and 100 µM of glucose plus 1 mM thiosulfate. Cell densities were monitored for growth by transferring 150 µL from each well to a matching 96 well polystyrene screening plate (Corning, Corning, NY). Syber Green I (Invitrogen, Carlsbad, CA)

was diluted in TRIS buffer and added to each well at a final concentration of 1/2000 and screened as previously described (Stingl *et al*., 2007). Pure cultures did not grow on any other media or in any other condition tested (autoclaved or boiled seawater, marine media 2216, R2A or 1/10 R2A media made with seawater, 1/10 R2A media in DSMZ #246, on agar plates, or in glass culture containers).

1. Blazejak A, Kuever J, Erseus C, Amann R, Dubilier N. (2006). Phylogeny of 16S rRNA, Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase, and Adenosine 5'-Phosphosulfate Reductase Genes from Gamma- and Alphaproteobacterial Symbionts in Gutless Marine Worms (Oligochaeta) from Bermuda and the Bahamas. Appl Environ Microbiol 72:5527-5536.

2. Canfield DE, Habicht KS, Thamdrup B. (2000) The Archean Sulfur Cycle and the Early History of Atmospheric Oxygen. Science 288:658-661.

3. Durand P, Benyagoub A, Prieur D. (1994) Numerical taxonomy of heterotrophic sulfur-oxidizing bacteria isolated from southwestern Pacific hydrothermal vents. Can J Microbiol 40:690-697.

4. Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J. (2005). Prokaryotic sulfur oxidation. Curr Opin Microbiol 8:253-259.

5. Hayes MK, Taylor GT, Astor Y, Scranton MI. (2006). Vertical distributions of thiosulfate and sulfite in the Cariaco Basin. Limnol Oceanogr 51:280–287.

6. Hipp WM, Pott AS, Thum-Schmitz N, Faath I, Dahl C, Truper HG. (1997). Towards the phylogeny of APS reductases ad sirohaem sulfite reductases in sulfate-reducing and sulfur-oxidizing prokaryotes. Microbiol 143:2891-2902.

7. Kelly DP, Shergill JK, Lu WP, Wood AP. (1997). Oxidative metabolism of inorganic sulfur compounds by bacteria. A Van Leeuw J 71:95-107.

8. Podgorsek L, Petri R, Imhoff JF. (2004) Cultured and genetic diversity, and activities of sulfur-oxidizing bacteria in low-temperature hydrothermal fluids of the North Fiji Basin. Mar Ecol Prog Ser 266: 65-76.

9. Ruby EG, Wirsen CO, Jannasch HW. (1981) Chemolithotrophic Sulfur-Oxidizing Bacteria from the Galapagos Rift Hydrothermal Vents. Appl Environ Microb 42:317-324.

10. Teske A, Brinkhoff T, Muyzer G, Moser DP, Rethmeier J, Jannasch HW. (2000). Diversity of Thiosulfate-Oxidizing Bacteria from Marine Sediments and Hydrothermal Vents. Appl Environ Microbiol 66: 3125-3133.

11. Tuttle JH, Holmes PE, Jannasch HW. (1974) Growth rate stimulation of marine pseudomonads by thiosulfate. Arch Microbiol 99:1-14.

12. Tuttle JH, Jannasch HW. (1976) Microbial utilization of thiosulfate in the deep sea. Limnol Oceanogr 21:697-701.

13. Tuttle JH, Jannasch HW. (1977) Thiosulfate stimulation of microbial dark assimilation of carbon dioxide in shallow marine environments. Microb Ecol 4:9-25.

14. Sievert SM, Kiene RP, Schultz-Vogt HN. (2007). The sulfur cycle. Oceanogr 20:117-123.

15. Sorokin DY. (2003). Oxidation of Inorganic Sulfur Compounds by Obligately Organotrophic Bacteria. Microbiol 72:641-653.

16. Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, *et al*. (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. Science 333:1296 -1300.

17. Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, *et al*. (2009). Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. Science 326:578-582.

18. Williams KP, Gillespie JJ, Sobral WS, Nordberg EK, Snyder EE, Shallom JM, Dickerman AW. (2010) Phylogeny of Gammaproteobacteria. J Bacteriol 192:2305-2314.

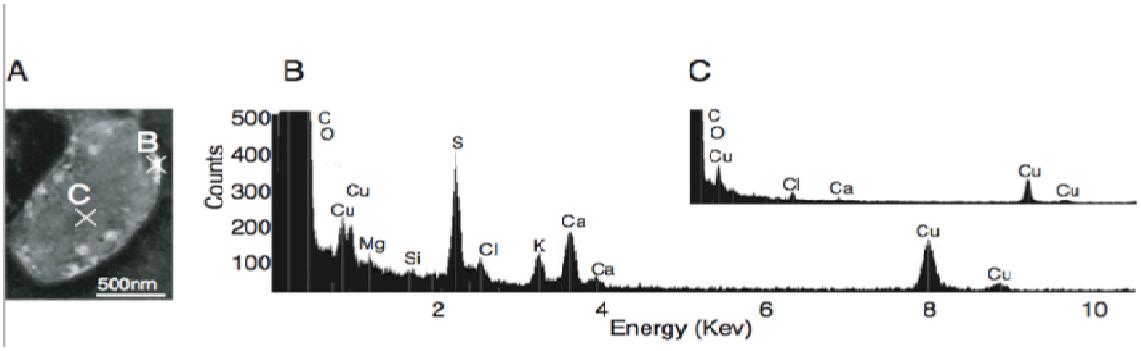

Figure 1.1. Image of *Candidatus* Thioglobus singularis and evidence of sulfur external globules determined by scanning transmission electron microscopy combined with energy dispersive x-ray spectroscopy (STEM-EDX). (A) STEM image and (B and C) EDX profiles. White extracellular globules contain sulfur (B) that was not detected in parts of the cell with no globule (C). Samples were analyzed on a copper (Cu) grid [1].

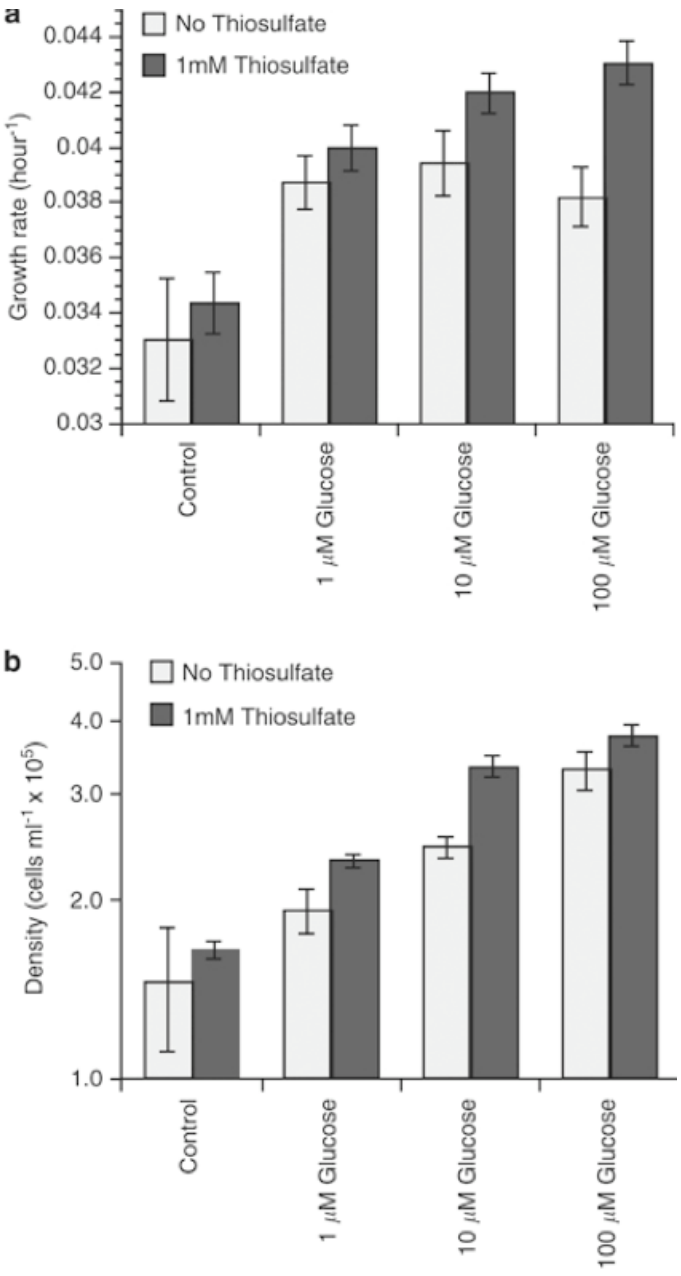

Figure 1.2. Growth rates and cell densities of *Candidatus* Thioglobus singularis. All experimental treatments were conducted in six replicate cultures and growth was plotted following initial detection (10^4 cells ml^{-1}). (A) Growth rates with glucose and glucose plus thiosulfate added to the media. (B) Final cell densities with glucose and glucose plus thiosulfate added to the media. Media was amended with 1 μM, 10 μM, or 100 μM glucose and each concentration of glucose plus 1 mM thiosulfate.

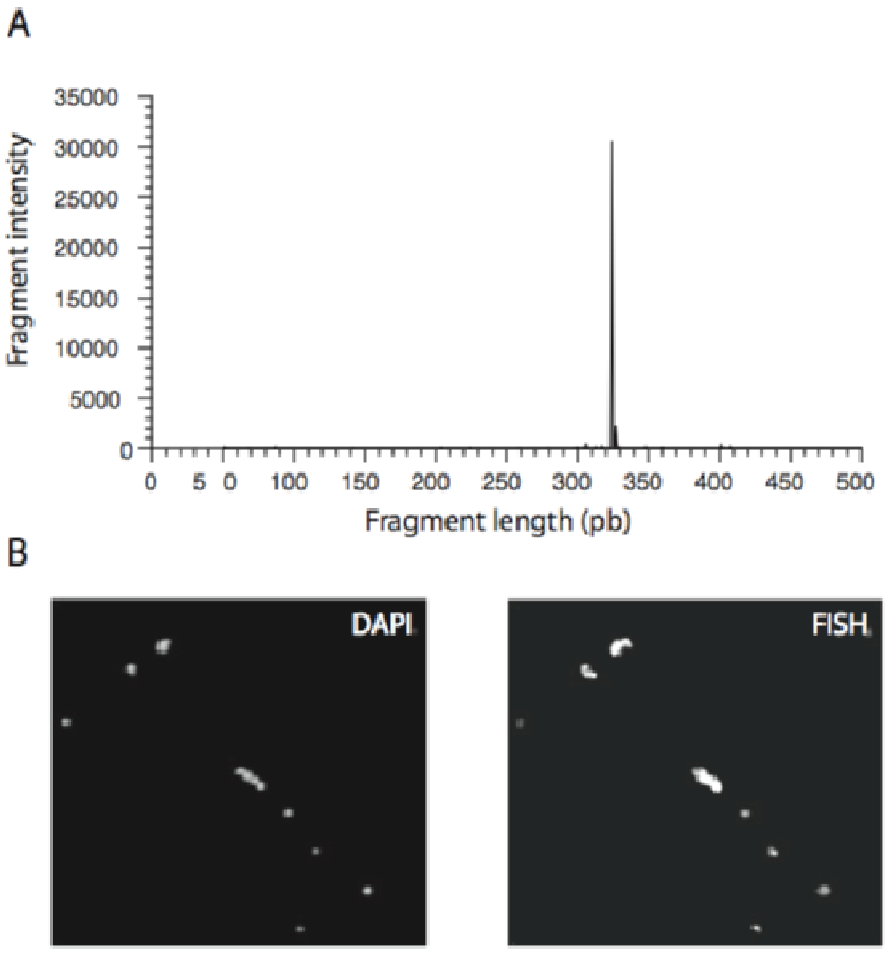

Figure 1.S1. Evidence of *T. singularis* GSO-PS1 purity. (A) A 324 bp TRFLP fragment matched the fragment size predicted from the 16S rRNA gene sequence. (B) Cells stained with the DNA stain DAPI and cells hybridized to the GSO-PS1 specific FISH probe (GSO-1448R). DAPI stains all cells that contain DNA and the GSO FISH probe 1448R is specific to Arctic96BD-19 cells. All DAPI stained cells had a positive GSO-1448R FISH signals. These data provide molecular and microscopic evidence confirming that we have pure cultures from the Arctic96BD-19 clade [1].

Table 1.S116S rRNA gene sequences of bacteria cultured from Puget Sound (PS) and from the North Pacific (NP).

Culture ID	Location	NCBI Accession	Bayesian Classification (p-values at each taxonimic level shown are > 0.7)
NPHTC-002	NP	JQ253970	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-020	NP	JQ253971	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-044	NP	JQ253972	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-026	NP	JQ253973	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-021	NP	JQ253974	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-043	NP	JQ253975	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
NPHTC-013	NP	JQ253976	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group
NPHTC-034	NP	JQ253977	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-027	NP	JQ253978	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group
NPHTC-040	NP	JQ253979	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-032	NP	JQ253980	Bacteria; Proteobacteria; Gammaproteobacteria; SUP05/Arctic96BD-19
NPHTC-005	NP	JQ253981	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group
NPHTC-028	NP	JQ253982	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-023	NP	JQ253983	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-035	NP	JQ253984	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group
NPHTC-012	NP	JQ253985	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group
NPHTC-024	NP	JQ253986	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-004	NP	JQ253987	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
NPHTC-029	NP	JQ253988	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group
NPHTC-042	NP	JQ253989	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-036	NP	JQ253990	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-018	NP	JQ253991	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group
NPHTC-016	NP	JQ253992	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-041	NP	JQ253993	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-007	NP	JQ253994	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-030	NP	JQ253995	Bacteria; Proteobacteria; Gammaproteobacteria; SUP05/Arctic96BD-19
NPHTC-011	NP	JQ253996	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-038	NP	JQ253997	Bacteria; Proteobacteria; Gammaproteobacteria; SUP05/Arctic96BD-19
NPHTC-014	NP	JQ253998	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-022	NP	JQ253999	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales; Rhodospirillaceae; Thalassospira
NPHTC-039	NP	JQ254000	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-008	NP	JQ254001	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-001	NP	JQ254002	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-015	NP	JQ254003	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-010	NP	JQ254004	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-019	NP	JQ254005	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-025	NP	JQ254006	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Sulfitobacter
NPHTC-033	NP	JQ254007	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-009	NP	JQ254008	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
NPHTC-017	NP	JQ254009	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
NPHTC-031	NP	JQ254010	Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales
NPHTC-003	NP	JQ254011	Bacteria; Proteobacteria; Gammaproteobacteria; SUP05/Arctic96BD-19
NPHTC-037	NP	JQ254012	Bacteria; Proteobacteria; Gammaproteobacteria; SUP05/Arctic96BD-19
NPHTC-006	NP	JQ254013	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Polaribacter
PSHTC-001	PS	JQ254014	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; environmental samples
PSHTC-002	PS	JQ254015	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae

PSHTC-003	PS	JQ254016	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-004	PS	JQ254017	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Erythrobacteraceae; Altererythrobacter
PSHTC-005	PS	JQ254018	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR11 cluster
PSHTC-006	PS	JQ254019	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Erythrobacteraceae; Altererythrobacter
PSHTC-007	PS	JQ254020	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group; SAR92 clade
PSHTC-008	PS	JQ254021	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR11 cluster
PSHTC-009	PS	JQ254022	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
PSHTC-010	PS	JQ254023	Bacteria; Proteobacteria; Gammaproteobacteria; SUP05/Arctic96BD-19
PSHTC-011	PS	JQ254024	Bacteria; Proteobacteria; Gammaproteobacteria; SUP05/Arctic96BD-19
PSHTC-012	PS	JQ254025	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-013	PS	JQ254026	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group; OM182 clade
PSHTC-014	PS	JQ254027	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group; OM182 clade
PSHTC-015	PS	JQ254028	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
PSHTC-016	PS	JQ254029	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-018	PS	JQ254030	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
PSHTC-019	PS	JQ254031	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR11 cluster
PSHTC-020	PS	JQ254032	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
PSHTC-021	PS	JQ254033	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR11 cluster
PSHTC-022	PS	JQ254034	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR11 cluster
PSHTC-023	PS	JQ254035	Bacteria; Proteobacteria; Gammaproteobacteria; SUP05/Arctic96BD-19
PSHTC-024	PS	JQ254036	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group
PSHTC-025	PS	JQ254037	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-026	PS	JQ254038	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Alteromonadaceae; Microbulbifer
PSHTC-027	PS	JQ254039	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-028	PS	JQ254040	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
PSHTC-029	PS	JQ254041	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR11 cluster
PSHTC-030	PS	JQ254042	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR11 cluster
PSHTC-031	PS	JQ254043	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; SAR11 cluster
PSHTC-032	PS	JQ254044	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group; OM60 clade
PSHTC-033	PS	JQ254045	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Erythrobacteraceae; Erythrobacter
PSHTC-034	PS	JQ254046	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
PSHTC-035	PS	JQ254047	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group; SAR92 clade
PSHTC-036	PS	JQ254048	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-037	PS	JQ254049	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
PSHTC-039	PS	JQ254050	Bacteria; Proteobacteria; Gammaproteobacteria; OMG group; SAR92 clade
PSHTC-040	PS	JQ254051	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-041	PS	JQ254052	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-042	PS	JQ254053	Bacteria; Proteobacteria; Alphaproteobacteria; SAR116 cluster
PSHTC-043	PS	JQ254054	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae
PSHTC-044	PS	JQ254055	Bacteria; Proteobacteria; Alphaproteobacteria; SAR116 cluster
PSHTC-045	PS	JQ254056	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-046	PS	JQ254057	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae
PSHTC-047	PS	JQ254058	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae

Chapter 2. A STREAMLINED MARINE MIXOTROPH

Katharine T. Marshall, Vaughn Iverson, E. Virginia Armbrust, Rhonda L. Morales, Chris T. Berthiaume, Bryndan Durham, Mary Ann Moran, Robert M. Morris (In progress)

2.1 ABSTRACT

The ocean's most abundant bacteria thrive in carbon and energy limited systems. They have adapted to these conditions in part by streamlining their genomes to reduce the amount of DNA they replicate and maintain to conserve energy and by using multiple metabolic strategies to meet carbon and energy demands — two strategies that are somewhat at odds with each other. The evolutionary pressure to maintain a small genome should limit metabolic flexibility and most pelagic marine bacteria have been characterized as primarily autotrophs that produce organic carbon or heterotrophs that consume exogenous organic carbon. Here we report the striking metabolic flexibility of a group of marine mixotrophs that have the genetic potential to function as both primary producers and primary consumers of dissolved organic carbon in seawater. Mixotrophic members of the SUP05 clade of sulfur-oxidizing marine bacteria maintain streamlined genomes, while retaining the metabolic potential to biosynthesize using organic and inorganic carbon and to produce energy using diverse organic and inorganic substrates. Genes to fix inorganic carbon via the CBB cycle and to transport and metabolize organic carbon compounds were expressed simultaneously by mixotrophic SUP05 cells in North Pacific surface waters, indicating that this group maintains and expresses core autotrophic and heterotrophic functions, despite evolutionary pressures to specialize in either autotrophy or

heterotrophy. Other bacteria may use this strategy as it takes advantage of multiple resources and SUP05 can be used as a model for learning about these streamlined mixotrophs.

2.2 SIGNIFICANCE STATEMENT

Here we report the striking metabolic flexibility of a ubiquitous group of streamlined mixotrophic marine bacteria, *Thioglobus singularis* sp. Comparative genomic analysis of the type strain, GSO-PS1, with environmental sequences from around the globe indicate that this group of marine bacteria can biosynthesize using organic or inorganic carbon as sole carbon sources and can produce energy using organic carbon, inorganic sulfur, and solar radiation. These highly conserved functions were detected in North Pacific surface water plankton communities, suggesting that genome reduction and metabolic flexibility are key evolutionary traits that are shaping microbial diversity and influencing ecosystem functions.

2.3 INTRODUCTION

Pelagic marine bacteria are well adapted to carbon and energy limited systems. Most have reduced the size of their genomes to conserve energy via an evolutionary process known as genome streamlining (1, 2, 3). Streamlined genomes have relatively few protein coding genes, few paralogous genes, and low GC content (1, 4). Genome streamlining is common in marine bacteria and occurs in environments where there are limited nutrients and large effective populations sizes (5, 6).

Mixotrophy is another strategy that can be helpful in environments with limited carbon and energy sources. It is typically used to describe an organism's ability to use organic and inorganic carbon for biosynthesis and is well known in marine microeukaryotes, including protists, ciliates, and flagellates (7-10). Pelagic ammonia oxidizing Achaea *Nitrosopumilus maritimus* have potential for complete mixotrophy (11-15). *N. maritimus* has a small genome

(1.64 Mbp) and the ability to fix carbon and transport and incorporate organic molecules. Other closely related Thaumarchaeota have different metabolic potentials. For example Thaumarchaeota from wastewater treatment plants were found to not fix carbon (16). This diversity in the metabolic strategy of closely related organisms suggests that metabolic capabilities are under strong evolutionary pressure.

A broader definition of mixotrophy includes describing an organisms' ability to use multiple sources of energy, such as organic and inorganic substrates, and light, here referred to as partial mixotrophy (17). Mixotrophy is particularly advantageous in environments where different forms of carbon and energy become limited so that the microbe does not depend on the presence of one particular molecule (17).

Genome streamlining and partial mixotrophy are common in marine bacteria. Two of the ocean's most abundant autotrophic and heterotrophic bacterial lineages have streamlined genomes, *Prochlorococcus and Pelagibacter*, respectively (1, 3, 18). *Pelagibacter* sp. are streamlined organoheterotrophs that use organic matter for carbon and energy, but are also phototrophs that can use light-driven proton pumps to supplement energy demands (19). *Prochlorococcus* sp. are streamlined photoautotrophs that use light energy to fix inorganic carbon via the Calvin-Benson-Bassham (CBB) cycle, but can also use some organic carbon for biosynthesis (20, 21). Some heterotrophic *Roseobacteria* sp. can use energy from phototrophy to fix a small amount of cellular carbon (3%), and can oxidize carbon monoxide or use light to supplement energy demands (22-24).

Members of the SUP05 clade are abundant and cosmopolitan in the oceans. They are found throughout the water column in oxic and anoxic ecosystems, and as symbionts in the tissues of deep-sea clams and mussels (25-32). They often co-occur, which has made it difficult

to identify functional groups with the SUP05 clade. We used the complete genome sequence of the first cultured representative from the SUP05 clade (*T. singularis* GSO-PS1) to elucidate the metabolic potential of a mixotrophic subclade within the SUP05 clade. Transcriptomic analysis of surface seawater plankton indicate that mixotrophic SUP05 cells in the environment express genes to fix inorganic carbon, to transport and assimilate dissolved organic carbon (DOC), and to use diverse sources of energy in the ocean surface layer. These data suggest that, similar to *Nitrosopumilus* Archaea, mixotrophs from the SUP05 bacterial clade have retained genes for autotrophy and heterotrophy, despite intense evolutionary pressures to specialize. These SUP05 can be used as model organisms to understand bacterial mixtrophy in pelagic waters.

2.4 Results and Discussion

T. singularis strain GSO-PS1, isolated from Puget Sound Washington, USA, contains key genes required for autotrophic and heterotrophic growth, including genes for inorganic carbon fixation, organic carbon uptake (transporters) and utilization (catabolic genes), sulfur oxidation, and phototrophy (Fig. 1). Core genes associated with these functions are also present in genomes constructed from environmental samples, including a *T. singularis* genome constructed from a Puget Sound metagenome, and six single cell genomes from the North and South Atlantic Ocean (Fig. 1A). Complete or nearly complete *T. singularis* genomes are surprisingly similar (Table 1). They are comparable in size (1.6-1.7 Mbp), number of protein coding genes (1.5-1.7 K), tRNAs (33-38), transporters (81-102), and other core metabolism genes (Table 1). The high number of transporters is key to organic carbon uptake for heterotrophy. Representative genomic regions around core genes for carbon fixation (RuBisCO) and sulfur oxidation (aprAB) indicate that genomic synteny is also highly conserved (Fig 1B and 1C) in *T. singularis* genomes from different oceans.

Though members of the *T. singularis* subclade are distributed throughout the water column, and are genetically similar, they can differ at the amino acid level with depth. *T. singularis* from surface waters in Puget Sound and in the North Atlantic are 92 to 99% percent identical to GSO-PS1 at the amino acid level, while *T. singularis* from 800 m in the South Atlantic are 72 to 75% identical to GSO-PS1 at the amino acid level (Table 1). This divergence suggests that there may be surface and deep ocean ecotypes. For example, proteorhodopsin (PR) genes used for phototrophy were not identified in either single cell genome from 800 meters depth in the South Atlantic. While this gene could have been missed because these are incomplete single cell genomes, it also could be absent from these organisms due to its uselessness without light at 800 meters. While Crenarchaea dominate carbon fixation in dark waters (33, 34), Swan and colleagues found that *T. singularis* were among the dominant Bacteria capable of fixing carbon below 200 m in the South Atlantic ocean (32), indicating that they are key players in a secondary source of primary production occurring in the deep ocean that is inextricably linked to the biological pump.

Comparative genomic analyses of *T. singularis* strain GSO-PS1 and a *T. autotrophica* metagenome (SI-SUP05) from Saanich Inlet, a seasonally anoxic fjord, revealed stark differences in the metabolic potential for members of their respective subclades to use carbon, nitrogen, and sulfur (Fig. 2 and Fig S1, Table 1). GSO-PS1 has enhanced genetic potential to use organic substrates in aerobic environments (Fig. 2A). Its genome codes for a complete TCA cycle and harbors key genes for glycolysis, gluconeogenesis, and the pentose phosphate pathway. All of these are missing from SI-SUP05 (Fig. S1, Fig. 1A and Table 1). These include the E1 and E2 components of the oxoglutarate dehydrogenase complex (OGDC), which is a key enzyme of the TCA cycle that is missing from SI-SUP05 and many other obligate autotrophs

(28, 35). GSO-PS1 also has 79 ABC transporters compared to only 23 ABC transporters in SI-SUP05 indicating the SI-SUP05 has limited ability to take up organic carbon. Genetic evidence for heterotrophy is confirmed by an increase in cell density when GSO-PS1 is grown with glucose (25).

T. singularis strain GSO-PS1 is missing key sulfide oxidation and dissimilatory nitrate reduction genes identified in SI-SUP05, though the genome does harbor genes which confer the ability to oxidize thiosulfate and sulfite (Fig. 2B). The presence of sulfur oxidizing *aprAB* and *sat* genes suggests that *T. singularis* cells have the potential to oxidize thiosulfate and sulfite via an adenosine 5'-phosphosulfate (APS) intermediate (36). Four rhodanese genes, two of which are closely related to rhodaneses in *T. autotrophica*, have suspected roles in sulfur oxidation. Rhodanese proteins have the common function of creating persulfide groups with a reactive sulfur (37). Although rhodaneses are not as well characterized in bacteria, Anantharaman and colleagues found that expression of rhodanese genes in *T. autotrophica* cells increased with other sulfur oxidation genes, suggesting that it is used to oxidize thiosulfate as part of the sulfur oxidation system (31). Although GSO-PS1 has four *dsr* genes (*dsrCMKL*) and produces sulfur globules (25), it lacks *dsrAB*, which are required in other organisms to oxidize elemental sulfur to sulfite. However, the sulfur oxidation pathways found in bacteria are diverse. Organisms often have genes involved in different pathways and often only partial pathways (36). Here we propose a pathway for GSO-PS1 sulfur oxidation that has been reduced through genome streamlining and may represent a streamlined and highly specialized sulfur oxidation system (Figure 2B).

T. singularis strain GSO-PS1 has genes to assimilate nitrogen in the form of ammonia or amino acids, but lacks nitrate transporters and other key nitrate assimilation genes. A putative *narI*, part of a membrane bound operon for nitrate reduction, was identified, but other respiratory

nitrate reduction genes are absent from the GSO-PS1 genome, including those identified in SI-SUP05 (Fig. 2C and Fig S1, Table 1) (26). GSO-PS1 does have a cytochrome C oxidase gene and a green-light adapted PR gene, indicating the potential to grow aerobically in sunlit surface waters. Differences in organic carbon assimilation and anaerobic and aerobic respiratory complexes suggest that *T. singularis* cells are mixotrophs that are adapted to aerobic environments and that *T. autotrophica* cells are sulfur-oxidizing chemolithoautotrophs that are adapted to hypoxic and anoxic environments.

To look for evidence of active and concurrent carbon fixation and heterotrophy we performed a genome-wide expression profile of *T. singularis* using metatranscripomic data collected from a North Pacific seawater community. Transcripts from surface waters (5m) at four stations along Line P were mapped to the *T. singularis* GSO-PS1 genome with high confidence (Fig. 3). Mean coverage of the GSO-PS1 genome was from 9.0X at station P1 (Std Dev 26.2 bp/gene) to 0.3X at P8 (Std Dev 1.8 bp/gene). The relative coverage of transcripts cannot be used to make biologically significant conclusions, only to show that more data exists from certain stations. At P1, of 1,714,149 nucleotides, 903,491 sites were identical, and pairwise identity was 93.5%. GSO-PS1 expressed genes for transporters for heterotrophic growth (77 of 79 ABC transporters—not pictured), carbon fixation (RuBisCO LS, *cbbOQR*), sulfur oxidation (*aprAB*, *sat*, four rhodaneses, *dsrCKLM*), and phototrophy (proteorhodopsin) (Fig. 3). This suggests that *T. singularis* cells were simultaneously functioning as inorganic carbon fixing autotrophs and organic carbon consuming heterotrophs in the ocean surface layer.

The *T. singularis* GSO-PS1 genome is similar in size, contains a similar number of protein coding genes, and a similar number of paralogous gene families to other pelagic marine bacteria (Fig 4, Fig. S2, Table 3). It closely resembles streamlined genomes from the ocean's most

abundant autotrophic and heterotrophic bacterial lineages, *Prochlorococcus* and *Pelagibacter*, which use partial mixotrophy (1, 3, 18), but is unique in its ability to use complete mixotrophy, fulfilling carbon and energy demands with organic and inorganic nutrients. This implies that *T. singularis* cells have dual roles in the global carbon cycle, producing and consuming DOC. The ability for *T. singularis* cells to function as both autotrophs and heterotrophs and to have streamlined genomes is unique among cultured Bacteria. However, the limited nutrients available in the pelagic ocean make this a viable survival strategy and GSO-PS1 may just be the first cultured representatives using a strategy that is more widely employed.

2.5 MATERIALS AND METHODS

Genome Sequencing - Cultures of *T. singularis* were grown in one-liter polycarbonate bottles of filter sterilized seawater media as previously described (25). Cells were then filtered onto sterile 0.2 µm polyethersulfone filters (Pall, Port Washington, NY), placed in 15 mL Teflon tubes containing 2 ml of sucrose lysis buffer (SLB), and flash frozen in liquid nitrogen. Cells were later lysed by adding 100 µL of 1mg/mL lysozyme and incubating at 4 °C for 60 minutes, then by adding 465 µL of 10% SDS and 250 µL of proteinase K and incubating at 55 °C for 2 hours. DNA was extracted and purified from cell lysates using DNeasy Blood and Tissue and Minelute kits according the manufacturers instructions, respectively (QIAGEN, Germantown, MD).

Genome Assembly - 20.35 µg of DNA was used to construct a mate pair library according to the SOLiD™v3.0 mate-pair protocol (Life Technologies, Foster City, CA). The *T. singularis* genome was assembled using SEAStAR v. 0.4.17 as previously described (38). The initial assembly was 97% complete and contained 177 gaps with a mean gap length of 198 bp. Gaps

were closed by PCR with custom primers designed using Geneious 7.0.4 (Biomatters, Auckland, NZ). Amplification products were visualized, gel purified, and sequenced by Genewiz (Foster City, CA). The complete genome sequence (1,714,149bp) was confirmed by performing additional PCR reactions to resolve any irregularities in the assembly and by comparing read coverage, physical coverage and insert sizes of mate pairs covering the genome (Fig S3, Fig 3). A Puget Sound *T. singularis* genome was assembled from a Puget Sound metagenome (10/10/2008) as previously described (38).

Genome Annotation - *T. singularis* genomes were annotated using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (39). Annotations were checked against RAST annotations (40, 41), IMG annotations (42), and in some cases by phylogenetic analyses to verify accuracy. Discrepancies were corrected and annotations were submitted to NCBI under the following accession numbers; *T. singularis* GSO-PS1, CP006911; *T. singularis* Puget Sound metagenome, PRJNA252014.

Gene Expression - Plankton (0.2-2.0 μm, 5 m depth) were collected at Station P1 (48°34.5 N, 125°30.0 W), P4 (48°39.0 N, 126°40.0 W), P6 (48°44.6 N, 127°40.0 W), and P8 (48°49.0 N, 128°40.0 W) in May 2012 by sequential filtration through a 2.0 and 0.2 μm pore-size filters (142 mm diameter polycarbonate and Supor filters, respectively) and with a McClain pump. Filters were flash frozen in liquid N_2 and stored at -80°C. RNA was prepared for sequencing as previously described (Satinsky et al 2014, Durham et al 2015). Briefly, cells were lysed in 10 mL of Ambion lysis buffer (AM8540G) + 0.5 mL each of 0.5 and 0.1 mm zirconia beads (BioSpec Products, Inc.). An 1,006 nt internal standard was added to the lysis buffer in known copy numbers (43). rRNAs were selectively removed using a subtractive hybridization method (44) with biotinylated rRNA probes specific to Bacterial and Archaeal 16S and 23S and eukaryotic

18S and 28S. rRNA-depleted samples were linearly amplified using the MessageAmp II-Bacteria kit (Ambion; Life Technologies). Amplified mRNAs were converted into cDNAs using the Superscript III First-Strand Synthesis System (Invitrogen; Life Technologies) followed by the NEBNext mRNA Second Strand Synthesis Module (New England Biolabs). cDNA was sheared ultrasonically to ~200-250 base pair fragments and HiSeq libraries were constructed for paired-end (2 x 150bp) sequencing (Illumina Inc., San Diego, CA) using the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA). Reads were joined using PANDAseq (45), and paired reads were trimmed using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). rRNA, tRNA, and internal standard sequences were removed using a BLASTn search against a database containing representative rRNA and tRNA sequences and internal standards (bit score cutoff ≥50).

T. singularis reads were identified from the P1, P4, P6, and P8 metatranscriptomes by performing a protein similarity search using RAPSearch2 (46) against reference sequences from the type strain, GSO-PS1. Reads with sequence identity scores of ≥80% and e-values <1e-20 were re-analyzed against the RefSeq protein database with GSO-PS1 added to confirm functional and taxonomic annotations. Only reads that had a hit to GSO-PS1 and reciprocal best hits to GSO-PS1 in RefSeq were selected for final analyses. Reads may have had equally good best hits to published *T. singularis* proteins available from single cell genomes in RefSeq, but not to proteins from other organisms. Transcripts are under NCBI accession number PRJNA272345.

Genome comparisons - GSO genomes were compared to the *T. singularis* PSI genome by using BRIG (47). The synteny of aprA and RuBisCo was determined using gene search on IMG then viewing neighborhood regions with the same top COG hit via top homolog. The synteny dot

plot of *T. singularis* to the SUP05 symbiont was created using the IMG tool DotPlot that uses Mummer (48).

Metabolic genes from complete *Prochlorococcus* and *Pelagibacter* genomes were compared to metabolic genes from complete and nearly-complete SUP05 and *T. singularis* genomes. Genes were categorized into autotrophy or heterotrophy (catabolic) using RAST and KEGG pathway categories, and into sulfur oxidation using RAST categories and Walsh et. al. (26). Genes of interest were marked present in genomes if they were annotated in the genome or if blasting the target gene against the genome gave a best hit with an E value of less than $1e^{-50}$.

2.6 Acknowledgements

We thank officers and crew of the *R/V Thompson* for help collecting environmental samples along Line P. We would also like to thank J. Saunders for help with phylogenetic analysis. This work was supported by grants from the National Science Foundation OCE-1205232 (A. Ingalls and R. Morris) and OCE-1232840 (R. Morris), and GRFP DGE-0718124 (K, Marshall).

2.7 Additional Information

Additional information that was not included in this publication can be found in Appendix B.

1. Giovannoni SJ, et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309(5738):1242–1245.

2. Mira A, Ochman H, Moran NA (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet* 17(10):589–596.

3. Rocap G, et al. (2003) Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature* 424(6952):1042–1047.

4. Swan BK, et al. (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci*.

5. Giovannoni SJ, Cameron Thrash J, Temperton B (2014) Implications of streamlining theory for microbial ecology. *ISME J* 8(8):1553–1565.

6. Dufresne A, Garczarek L, Partensky F (2005) Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol* 6(2):R14.

7. Riemann B, Havskum H, Thingstad F, Bernard C (1995) The Role of Mixotrophy in Pelagic Environments. *Molecular Ecology of Aquatic Microbes*, NATO ASI Series., ed Joint I(Springer, Berlin Heidelberg), pp 87–114.

8. Stoecker DK (1998) Conceptual models of mixotrophy in planktonic protists and some ecological and evolutionary implications. *Eur J Protistol* 34(3):281–290.

9. Jones RI (2000) Mixotrophy in planktonic protists: an overview. *Freshw Biol* 45(2):219–226.

10. Sanders RW (1991) Mixotrophic Protists In Marine and Freshwater Ecosystems. *J Protozool* 38(1):76–81.

11. Ingalls AE, et al. (2006) Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. *Proc Natl Acad Sci* 103(17):6442–6447.

12. Walker CB, et al. (2010) Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci* 107(19):8818–8823.

13. Kuypers MMM, et al. (2001) Massive Expansion of Marine Archaea During a Mid-Cretaceous Oceanic Anoxic Event. *Science* 293(5527):92–95.

14. Ouverney CC, Fuhrman JA (2000) Marine Planktonic Archaea Take Up Amino Acids. *Appl Environ Microbiol* 66(11):4829–4833.

15. Pearson A, McNichol AP, Benitez-Nelson BC, Hayes JM, Eglinton TI (2001) Origins of lipid biomarkers in Santa Monica Basin surface sediment: a case study using compound-specific Δ14C analysis. *Geochim Cosmochim Acta* 65(18):3123–3137.

16. Mußmann M, et al. (2011) Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci* 108(40):16771–16776.

17. Eiler A (2006) Evidence for the ubiquity of mixotrophic bacteria in the upper ocean: implications and consequences. *Appl Environ Microbiol* 72(12):7431–7437.

18. Strehl B, Holtzendorff J, Partensky F, Hess WR (1999) A small and compact genome in the marine cyanobacterium *Prochlorococcus marinus* CCMP 1375: lack of an intron in the gene for tRNA(Leu)UAA and a single copy of the rRNA operon. *FEMS Microbiol Lett* 181(2):261–266.

19. Giovannoni SJ, et al. (2005) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* 438(7064):82–85.

20. Zubkov MV, Fuchs BM, Tarran GA, Burkill PH, Amann R (2003) High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl Environ Microbiol* 69(2):1299–1304.

21. Zubkov MV, Tarran GA (2005) Amino acid uptake of *Prochlorococcus* spp. in surface waters across the South Atlantic Subtropical Front. *Aquat Microb Ecol* 40(3):241–249.

22. Kolber ZS (2001) Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* 292(5526):2492–2495.

23. Moran MA, et al. (2007) Ecological genomics of marine *Roseobacters*. *Appl Environ Microbiol* 73(14):4559–4569.

24. Yurkov VV, Beatty JT (1998) Aerobic Anoxygenic Phototrophic Bacteria. *Microbiol Mol Biol Rev* 62(3):695–724.

25. Marshall KT, Morris RM (2013) Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J* 7(2):452–455.

26. Walsh DA, et al. (2009) Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 326(5952):578–582.

27. Anderson RE, Beltrán MT, Hallam SJ, Baross JA (2013) Microbial community structure across fluid gradients in the Juan de Fuca Ridge hydrothermal system. *FEMS Microbiol Ecol* 83(2):324–339.

28. Zaikova E, et al. (2010) Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* 12(1):172–191.

29. Anantharaman K, Breier JA, Sheik CS, Dick GJ (2012) Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proc Natl Acad Sci* 110(1):330–335.

30. Lavik G, et al. (2009) Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* 457(7229):581–584.

31. Sunamura M, Higashi Y, Miyako C, Ishibashi J, Maruyama A (2004) Two bacteria phylotypes are predominant in the Suiyo Seamount hydrothermal plume. *Appl Env Microbiol* 70(2):1190–1198.

32. Swan BK, et al. (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333(6047):1296 –1300.

33. Berg IA, Kockelkorn D, Buckel W, Fuchs G (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 318(5857):1782–1786.

34. Wuchter C, et al. (2006) Archaeal nitrification in the ocean. *Proc Natl Acad Sci* 103(33):12317–12322.

35. Smith AJ, London J, Stanier RY (1967) Biochemical basis of obligate autotrophy in blue-green algae and *Thiobacilli*. *J Bacteriol* 94(4):972–983.

36. Ghosh W, Dam B (2009) Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. *FEMS Microbiol Rev* 33(6):999–1043.

37. Cipollone R, Ascenzi P, Visca P (2007) Common themes and variations in the rhodanese superfamily. *IUBMB Life* 59(2):51–59.

38. Iverson V, et al. (2012) Untangling genomes from metagenomes: revealing an uncultured class of marine euryarchaeota. *Science* 335(6068):587–590.

39. Angiuoli SV, et al. (2008) Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *Omics J Integr Biol* 12(2):137–141.

40. Aziz RK, et al. (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75.

41. Overbeek R, et al. (2013) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206-D214.

42. Markowitz VM, et al. (2012) IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 40:D115–D122.

43. Satinsky BM, Gifford SM, Crump BC, Moran MA (2013) Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods in Enzymology*, ed DeLong EF (Elsevier Inc.), pp 237–250. 1st Ed.

44. Stewart FJ, Ottesen EA, DeLong EF (2010) Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J* 4(7):896–907.

45. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD (2012) PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 13:31.

46. Zhao Y, Tang H, Ye Y (2012) RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* 28(1):125–126.

47. Alikhan N-F, Petty NK, Zakour NLB, Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12(1):402.

48. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5(2):R12.

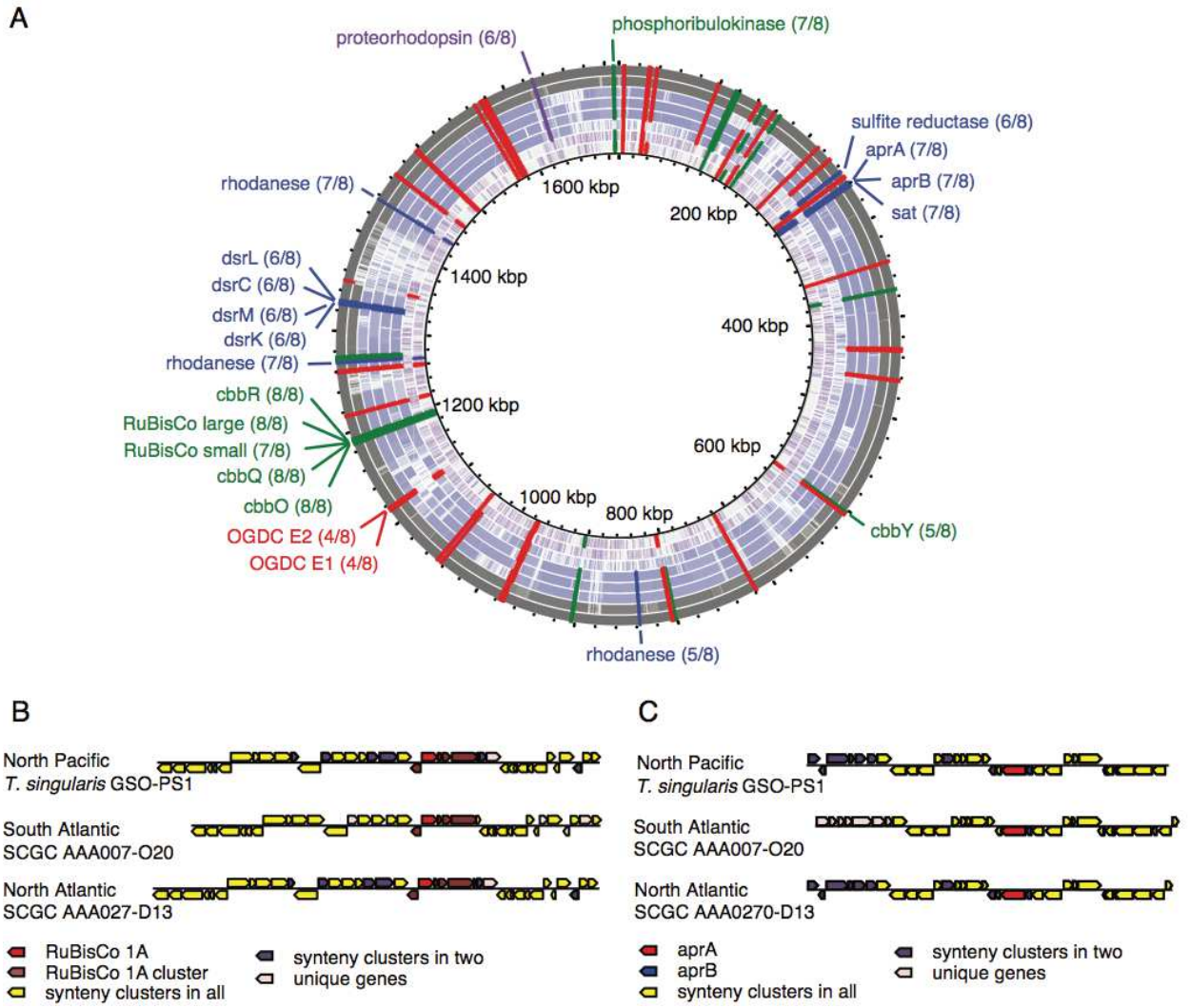

Figure 2.1. Comparative analysis of *T. singularis* genomes. A) Shared genes associated with mixotrophic carbon and energy metabolisms. Rings from outermost to innermost: GSO-PS1 type strain; a genome constructed from a Puget Sound metagenome; and six single cell genomes—from the surface North Atlantic: AAA076-D13, AAA076-D02, AAA076-F14, AAA076-E13—from the deep South Atlantic: AAA007-B15, AAA007-O20, respectively. Shaded areas correspond to genes with amino acid identity scores >80%. Markers indicate key genes for autotrophic (green), heterotrophic (catabolic only, transporters not pictured) (red), lithotrophic (blue), and phototrophic (purple) growth. Catabolic genes include those involved in the TCA cycle, pentose phosphate pathway, glycolysis, and gluconeogenesis. B and C) Representative regions of synteny associated with carbon fixation (RuBisCo cluster) and sulfur oxidation (aprA/aprB). Gene abbreviations: dsr – dissimilatory sulfite reductase; apr – adenosine phosphosulfate reductase; sat – ATP sulfurylase; OGDC – oxoglutarate dehydrogenase complex; cbb – Calvin-Benson-Bassham cycle; RuBisCO – Ribulose-1,5-bisphosphate carboxylase/oxygenase.
[1].

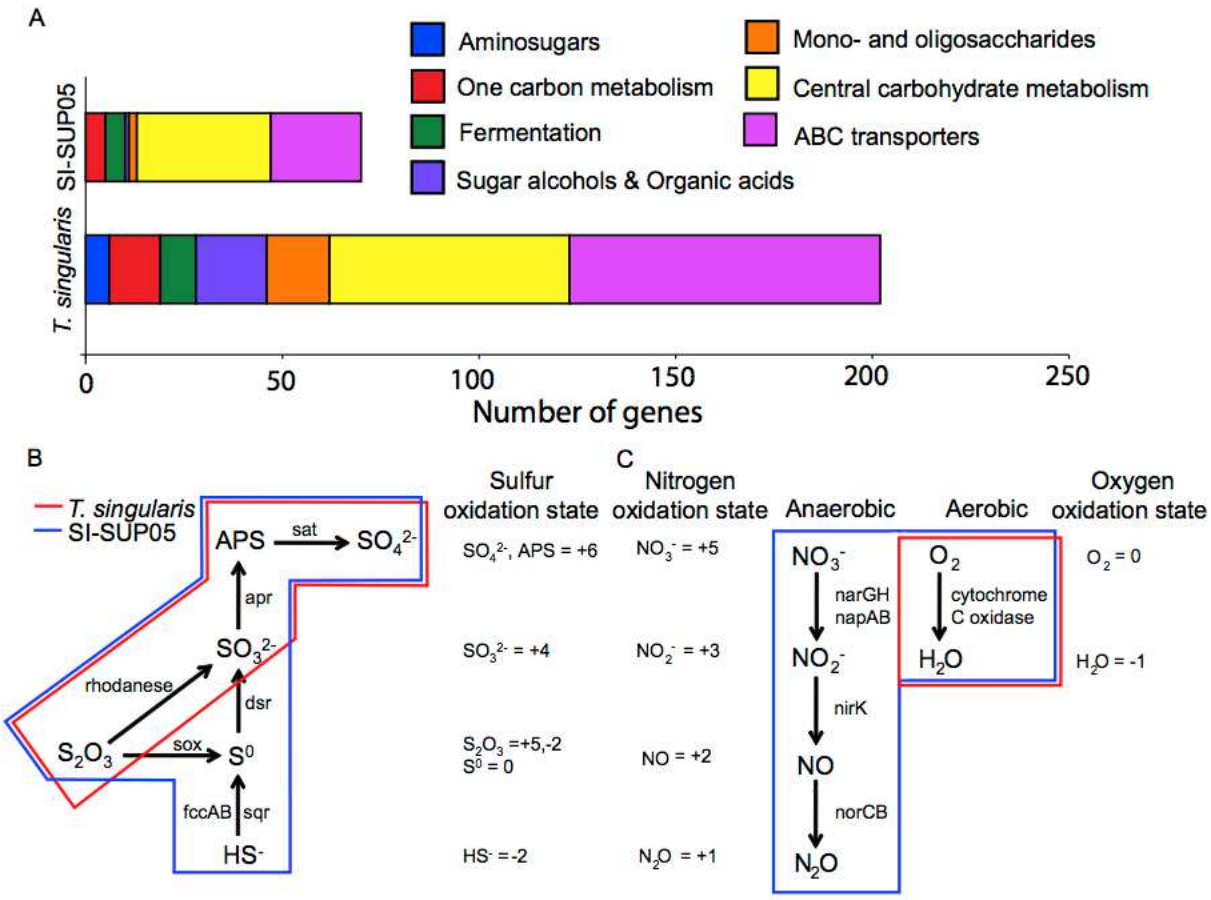

Figure 2.2. Comparison of carbon assimilation and respiratory genes in *T. singularis* GSO-PS1 and *T. autotrophica* SI-SUP05. A) Number of genes associated with carbohydrate metabolisms. B) Genes associated with sulfur oxidation. Question marks indicate unknown genes that are expected given the elemental sulfur globules on *T. singularis* (Marshall and Morris, 2013) C) Genes associated with respiration under aerobic and anaerobic conditions. Gene abbreviations: dsr – dissimilatory sulfite reductase; sox – sulfur oxidation system; apr – adenosine phosphosulfate reductase; sat – ATP sulfurylase; sqr – sulfide quinone reductase; nar – nitrate reductase; nir – nitrite reductase; nor – nitric oxide reductase [1].

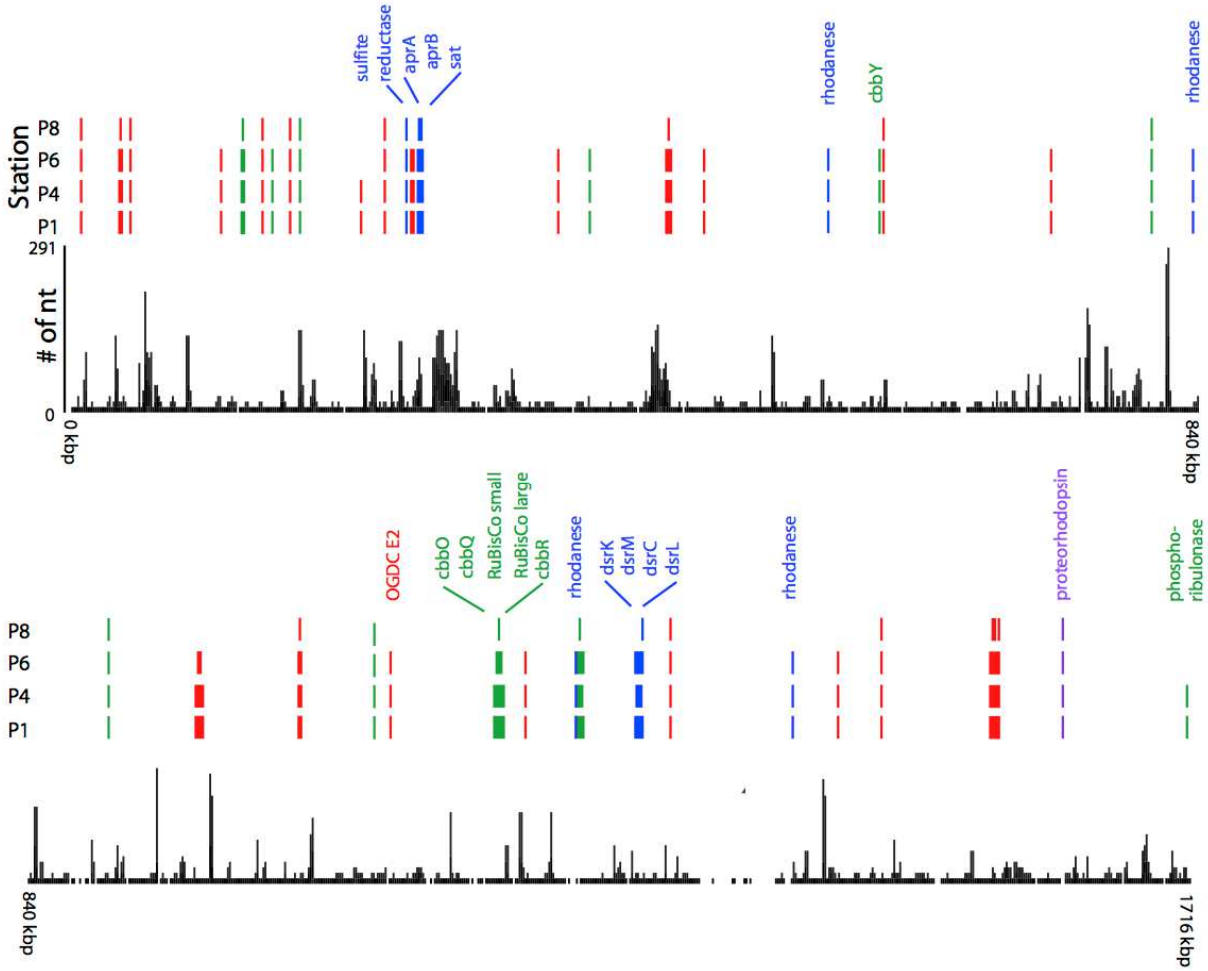

Figure 2.3. Transcript coverage of *T. singularis* GSO-PS1 genome. Transcripts from four stations along Line P in the North Pacific were aligned to the GSO-PS1 genome using Geneious 7.0.4 (Biomatters, Auckland, NZ). P1 had the deepest sequencing and provided the highest transcript coverage (mean coverage 9.0X) of the *T. singularis* genome (black line). Its pairwise identity was 93.5%. Markers indicate transcripts identified for autotrophic (green), catabolic (heterotrophic) (red), lithotrophic – sulfur oxidation (blue), and phototrophic – proteorhodopsin (purple) metabolisms (as in Figure 1). Catabolic transcripts that were identified include those involved in the TCA cycle, pentose phosphate pathway, glycolysis, and gluconeogenesis. Gene abbreviations: *dsr* – dissimilatory sulfite reductase; *apr* – adenosine phosphosulfate reductase; sat – ATP sulfurylase; OGDC – oxoglutarate dehydrogenase complex; cbb – Calvin-Benson-Bassham cycle; RuBisCO – Ribulose-1,5-bisphosphate carboxylase/oxygenase [1].

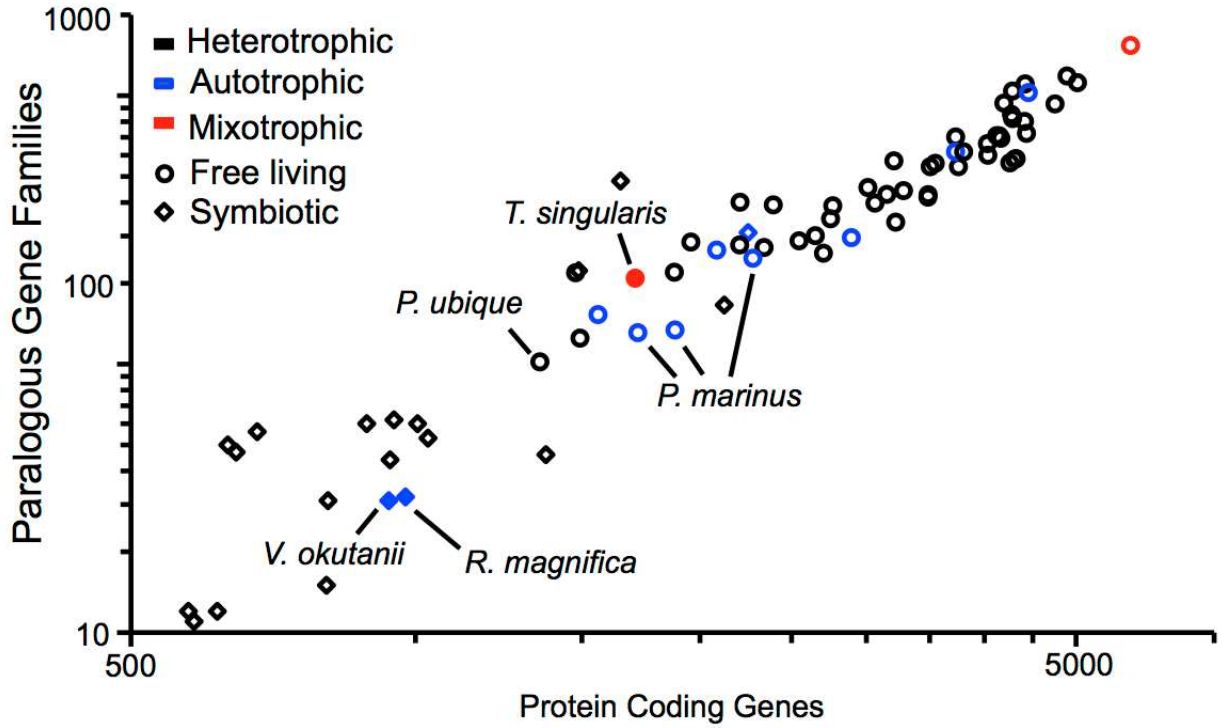

Figure 2.4. Streamlined *T. singularis* GSO-PS1 genome. Number of protein coding genes versus paralogous gene families [1].

Table 2.1 Comparison of *T. singularis* genomes. The complete genome of *T. singularis* strain GSO-PS1 was compared to a *T. singularis* genome assembled from Puget Sound and six partial *T. singularis* single cell genomes. Four single cell genomes derived from the North Atlantic (AAA076) and two derived from the South Atlantic (AAA001 and AAA007).

	GSO PS1	*Puget Sound	AAA076 F14	AAA076 **E13	AAA076 D02	AAA076 D13	AAA007 **O20	AAA001 **B15
Genome size (bp)	1,714,148	1,635,114	1,763,976	953,811	1,761,763	1,663,375	1,028,107	771,017
Mean AA % identity		99.1	95.4	92.2	95.4	95.7	74.7	72.1
Protein coding genes	1,750	1,574	1742	903	1739	1685	966	803
% GC	37.4	38.2	37.2	36.7	37.4	37.4	38	36.8
tRNAs	38	38	33	18	36	34	17	12
rRNAs	3	3	3	3	3	3	2	3
CO2 fixation genes	19	19	18	12	19	17	11	8
TCA cycle genes	13	13	11	7	11	13	7	3
ABC transporters	79	78	99	52	93	92	46	16
TRAP transporters	3	3	3	0	5	5	1	3
Sulfur metabolism	11	11	11	6	11	11	5	4
Proteorhodopsin	1	1	1	1	1	1	0	0

*A Puget Sound *T. singularis* genome was assembled using published data and as described by Iverson and colleagues (Iverson et al., 2012).

**Genomes that are less complete based on the number of core housekeeping genes identified in *T. singularis* and other bacteria. Theses include tRNAs, carbon fixation genes, TCA cycle genes, and sulfur metabolism genes.

Table 2.2 Genomic comparison of carbon, nitrogen and sulfur metabolisms in free-living *T. singularis* strain GSO-PS1 and Saanich Inlet SUP05 (SI-SUP05) cells. Genes were identified by comparing annotations and by confirmation via BLAST analyses with E values <1e-50.

	GSO-PS1	SI Metagenome
Carbon fixation		
RuBisCo Small	X	
RuBisCo Large	X	X
Rubisco operon transcriptional regulator (cbbR)	X	X
CbbQ	X	X
cbbO	X	X
cbbY	X	
phosphoglycolate phosphatase	X	X
serine hydroxymethyltransferase	X	X
Glycerate kinase	X	
glycine dehydrogenase	X	
glycine cleavage system H protein	X	X
Aminomethyltransferase	X	X
glycolate dehydrogenaseD	X	
glycolate dehydrogenaseE	X	X
glycolate dehydrogenaseF	X	X
alanine: glyoxylate aminotransferase	X	
2-hydroxy-3-oxopropionate reductase	X	
Malate synthase G	X	
phosphoribulokinase	X	
Central Carbohydrate Metabolism		
glucokinase	X	
6-phosphofructokinase	X	X
fructose-bisphosphatase		X
fructose-bisphosphate aldolase	X	X
triosephosphate isomerase	X	
glyceraldehyde-3-phosphate dehydrogenase	X	X
phosphoglycerate kinase	X	X
phosphoglyceromutase	X	X
phosphoglucomutase	X	X
enolase	X	X
pyruvate kinase	X	X
phosphoenolpyruvate synthase	X	
phosphoenolpyruvate carboxykinase	X	
aldose 1-epimerase	X	
citrate synthase	X	X
aconitase	X	X
2-oxoglutarate dehydrogenase E1	X	
2-oxoglutarate dehydrogenase E2	X	
succinyl-CoA synthetase subunit alpha	X	
succinyl-CoA synthetase subunit beta	X	
succinate dehydrogenase iron-sulfur protein	X	X
succinate dehydrogenase flavoprotein subunit	X	X
fumarate hydratase	X	X
malate dehydrogenase	X	X
isocitrate dehydrogenase	X	
dihydrolipoamide dehydrogenase	X	X

Dihydrolipoamide dehydrogenase		
glucose-6-phosphate 1-dehydrogenase	X	
6-phosphogluconolactonase	X	
6-phosphogluconate dehydrogenase	X	
ribose 5-phosphate isomerase	X	X
ribulose-phosphate 3-epimerase	X	X
transketolase	X	X
ribose-phosphate pyrophosphokinase	X	X
phosphopentomutase	X	
deoxyribose-phosphate aldolase	X	
gluconate kinase	X	
Methylglyoxal Metabolism		
Lactoylglutathione lyase	X	X
Aldehyde dehydrogenase B	X	
Aldehyde dehydrogenase	X	X
Hydroxyacylglutathione hydrolase	X	
Dihydrolipoamide dehydrogenase of pyruvate dehydrogenase	X	X
Dihydrolipoamide acetyltransferase of pyruvate dehydrogenase	X	X
Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate		
Pyruvate dehydrogenase E1 component	X	X
Phosphate acetyltransferase	X	
Acetyl-coenzyme A synthetase	X	
Acetyl-CoA synthetase (ADP-forming) alpha and beta chains	X	
Acylphosphate phosphohydrolase		X
Pyruvate Alanine Serine Interconversions		
Alanine dehydrogenase	X	X
Alanine racemase	X	X
L-serine dehydratase	X	
D-alanine aminotransferase	X	X
Omega-amino acid--pyruvate aminotransferase	X	
D-amino acid dehydrogenase small subunit	X	
Branched-chain amino acid aminotransferase	X	
Glyoxylate bypass		
Isocitrate dehydrogenase phosphatase/kinase	X	
Subsystem: Dehydrogenase complexes		
Cytosol aminopeptidase PepA	X	X
Pyruvate metabolism I: anaplerotic reactions, PEP		
NADP-dependent malic enzyme	X	X
Oxaloacetate decarboxylase beta chain	X	
Oxaloacetate decarboxylase alpha chain	X	
Glycolate, glyoxylate interconversions		
phosphoglycolate phosphatase	X	X
glycolate dehydrogenaseD	X	
glycolate dehydrogenaseE	X	X
glycolate dehydrogenaseF	X	X

	GSO-PS1	SI Metagenome
Amino sugars		
N-acetylglucosamine-6-phosphate deacetylase	X	
Glucosamine-6-phosphate deaminase	X	
N-Acetyl-D-glucosamine ABC transport system	X	
N-Acetyl-D-glucosamine ABC transport system	X	
Transcriptional regulator of N-Acetylglucosamine	X	
Beta-hexosaminidase	X	
One carbon metabolism		
Malyl-CoA lyase	X	
Serine-pyruvate aminotransferase	X	
Methylenetetrahydrofolate dehydrogenase	X	X
Formate--tetrahydrofolate ligase	X	
5,10-methylenetetrahydrofolate reductase	X	
Acetyl-CoA acetyltransferase	X	X
Isocitrate lyase	X	
Succinyl-CoA ligase [ADP-forming] alpha chain	X	
Succinyl-CoA ligase [ADP-forming] beta chain	X	
5-formyltetrahydrofolate cyclo-ligase	X	X
Formyltetrahydrofolate deformylase		X
serine hydroxymethyltransferase	X	X
L-alanine:glyoxylate aminotransferase	X	
Glycerate kinase	X	
Fermentation		
Butyryl-CoA dehydrogenase	X	X
Enoyl-CoA hydratase	X	X
Alcohol dehydrogenase	X	
Acetolactate synthase large subunit	X	X
Acetolactate synthase small subunit	X	X
3-hydroxyacyl-CoA dehydrogenase	X	X
Electron transfer flavoprotein, alpha subunit	X	
Electron transfer flavoprotein, beta subunit	X	
Electron transfer flavoprotein-ubiquinone oxidoreductase	X	
Sugar Alcohols and Organic acids		
Glycerol-3-phosphate regulon repressor, DeoR family	X	
Glycerol kinase	X	
Aerobic glycerol-3-phosphate dehydrogenase	X	
Glycerol-3-phosphate dehydrogenase [NAD(P)+]	X	X
Glycerophosphoryl diester phosphodiesterase	X	
Glycerol-3-phosphate ABC transporter, ATP-binding protein UgpC	X	
Glycerol-3-phosphate ABC transporter, glycerol-3-phosphate-binding	X	
Myo-inositol 2-dehydrogenase	X	
Inosose dehydratase	X	
Epi-inositol hydrolase	X	
5-deoxy-glucuronate isomerase	X	
5-keto-2-deoxygluconokinase	X	
5-keto-2-deoxy-D-gluconate-6 phosphate aldolase	X	
Predicted transcriptional regulator of the myo-inositol	X	
Inositol transport system permease protein	X	
Inositol transport system sugar-binding protein	X	

Inositol transport system ATP-binding protein	X	
2-hydroxy-3-oxopropionate reductase	X	
	GSO-PS1	SI Metagenome
Mono and Oligo-saccharides		
Ribokinase	X	
Ribose 5-phosphate isomerase A	X	X
Predicted nucleoside ABC transporter, permease 1 component	X	
Predicted nucleoside ABC transporter, permease 2 component	X	
Predicted nucleoside ABC transporter, ATP-binding component	X	
Predicted nucleoside ABC transporter, substrate-binding component	X	
Galactonate dehydratase	X	
2-dehydro-3-deoxygalactonokinase	X	
2-dehydro-3-deoxyphosphogalactonate aldolase	X	
Thymidine phosphorylase	X	
Purine nucleoside phosphorylase	X	
Deoxyribose-phosphate aldolase	X	
Putative deoxyribonuclease YjjV	X	X
Putative deoxyribose-specific ABC transporter, permease protein	X	
6-phosphogluconate dehydrogenase, decarboxylating	X	
Beta-galactosidase	X	
Transporters		
ABC transporters	79	23
TRAP transporters	3	0
Sulfur Oxidation		
aprA	X	X
aprB	X	
dsrA		X
dsrB		X
dsrC	X	X
dsrE		X
dsrF		X
dsrH		X
dsrK	X	X
dsrL	X	
dsrM	X	X
fccA		X
rhodanese	X	X
rhodanese	X	X
rhodanese	X	
sat	X	X
soxA		X
soxB		X
soxF		X
soxH		X

soxW		X
soxX		X
soxY		X
soxZ		X
sqr		X
sulfite reductase	X	
	GSO-PS1	SI Metagenome
Denitrification		
narK		X
$narK_2$		X
narG		X
narH		X
narJ		X
narI	X	X
napF		X
napB		X
napA		X
napH		X
napG		X
napD		X
nirK		X
norC		X
norB		X
mobA	X	X

Table 2.3. Seventy-six complete bacterial genomes were compared to the GSO-PS1 genome to evaluate streamlining. Genomes selected for streamlining analyses included symbionts and bacteria from aquatic environments. Genomes were evaluated as previously described by Giovanonni, et. al., 2005 (1).

Organism ID	NCBI Accession #
T. singularis GSO-PS1	
Alcanivorax borkumensis SK2	GCA_000009365.1
Alcanivorax dieselolei B5	GCA_000300005.1
Alpha proteobacterium HIMB59	GCA_000299115.1
Alteromonas macleodii AltDE1	GCA_000310085.1
Anabaena variabilis ATCC 29413	GCA_000204075.1
Anaplasma marginale str. St. Maries	GCA_000011945.1
Aquifex aeolicus VF5	GCA_000008625.1
Bartonella henselae strain Houston-1	GCA_000046705.1
Borrelia burgdorferi B31	GCA_000008685.2
Buchnera aphidicola str. APS	GCA_000009605.1
Candidatus Blochmannia floridanus	GCA_000043285.1
Candidatus Pelagibacter ubique HTCC1062	GCA_000012345.1
Candidatus Puniceispirillum marinum IMCC1322	GCA_000024465.1
Candidatus Ruthia magnifica str. Cm	GCA_000015105.1
Candidatus Vesicomyosocius okutanii HA	GCA_000010405.1
Cellulophaga algicola DSM 14237	GCA_000186265.1
Cellulophaga lytica DSM 7489	GCA_000190595.1
Chlamydia muridarum Nigg	GCA_000006685.1
Chlamydophila caviae GPIC	GCA_000007605.1
Chlorobium luteolum DSM 273	GCA_000012485.1
Coxiella burnetii RSA 493	GCA_000007765.1
Croceibacter atlanticus HTCC2559	GCA_000196315.1
Dehalococcoides mccartyi DCMB5	GCA_000341655.1
Dinoroseobacter shibae DFL 12	GCA_000018145.1
Ehrlichia ruminantium strain Welgevonden	GCA_000026005.1
Endosymbiont of unidentified scaly snail isolate Monju	GCA_000801295.1
Erythrobacter litoralis HTCC2594	GCA_000013005.1
Flavobacterium johnsoniae UW101	GCA_000016645.1
Fluviicola taffensis DSM 16823	GCA_000194605.1
Hirschia baltica ATCC 49814	GCA_000023785.1
Hyphomonas neptunium ATCC 15444	GCA_000013025.1
Idiomarina loihiensis GSL 199	GCA_000401175.1
Jannaschia sp. CCS1	GCA_000013565.1
Maricaulis maris MCS10	GCA_000014745.1
Marinithermus hydrothermalis DSM 14884	GCA_000195335.1

Marinobacter adhaerens HP15	GCA_000166295.1
Marinobacter aquaeolei VT8	GCA_000015365.
Marinomonas mediterranea MMB-1	GCA_000192865.1
Mesoplasma florum L1	GCA_000008305.1
Methylotenera mobilis JLW8	GCA_000023705.1
Methylotenera versatilis 301	GCA_000093025.1
Mycoplasma mobile 163K	GCA_000008365.1
Oceanithermus profundus DSM 14977	GCA_000183745.1
Octadecabacter antarcticus 307	GCA_000155675.2
Owenweeksia hongkongensis DSM 17368	GCA_000236705.1
Phaeobacter gallaeciensis 2.10	GCA_000154745.2
Planctomyces brasiliensis DSM 5305	GCA_000165715.3
Planctomyces limnophilus DSM 3776	GCA_000092105.1
Prochlorococcus marinus MIT9313	GCA_000011485.1
Prochlorococcus marinus subsProchlorococcus str. CCMP1375	GCA_000007925.1
Prochlorococcus marinus MED4	GCA_000011465.1
Pseudoalteromonas haloplanktis str. TAC125	GCA_000026085.1
Psychromonas ingrahamii 37	GCA_000015285.1
Rickettsia conorii str. Malish 7	GCA_000007025.1
Roseobacter denitrificans OCh 114	GCA_000014045.1
Ruegeria pomeroyi DSS-3	GCA_000011965.2
Shewanella denitrificans OS217	GCA_000013765.1
Shewanella frigidimarina	GCA_000014705.1
Sphingopyxis alaskensis RB2256	GCA_000013985.1
Spirochaeta thermophila DSM 6192	GCA_000147075.1
Synechococcus sp. CC9311	GCA_000014585.1
Teredinibacter turnerae T7901	GCA_000023025.1
Thermosipho africanus TCF52B	GCA_000021285.1
Thermosipho melanesiensis BI429	GCA_000016905.1
Thioflavicoccus mobilis 8321	GCA_000327045.1
Treponema pallidum subsp. pallidum str. Nichols	GCA_000008605.1
Trichodesmium erythraeum IMS101	GCA_000014265.1
Tropheryma whipplei str. Twist	GCA_000007485.1
Ureaplasma urealyticum serovar 10 str. ATCC 33699	GCA_000021265.1
Vibrio cholerae LMA3894-4	GCA_000195065.1
Vibrio fischeri ES114	GCA_000011805.1
Vibrio splendidus LGP32	GCA_000091465.1
Wigglesworthia glossinidia endosymbiont	GCA_000008885.1
Wolbachia endosymbiont strain TRS	GCA_000008385.1
Acaryochloris marina MBIC11017	GCA_000018105.1
Haliangium ochraceum DSM 14365	GCA_000024805.1

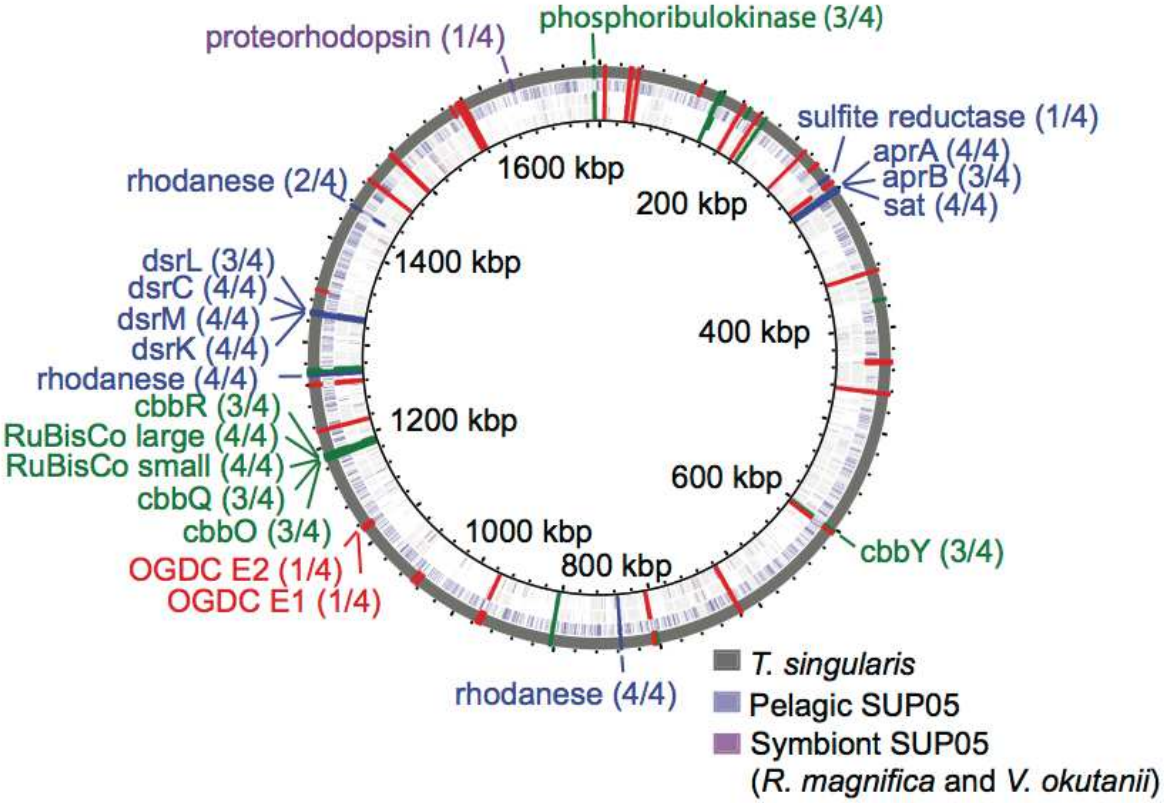

Figure 2.S1. Comparative analysis of *T. singularis* GSO-PS1 and autotrophic SUP05 subclade genomes. Genes indicating the potential for mixotrophic carbon and energy metabolisms. Rings from outermost to innermost; GSO-PS1 type strain; a genome constructed from a Saanich Inlet metagenome (SI-SUP05); and two symbiont genomes - *R. magnifica* and *V. okutanii*, respectively. Shaded areas in concentric rings correspond to genes with amino acid identity scores >80%. Markers indicate key genes for autotrophy (green), catabolic (heterotrophy) (red), lithotrophy (blue), and phototrophy (purple). Catabolic genes include those involved in the TCA cycle, pentose phosphate pathway, glycolysis, and gluconeogenesis [1].

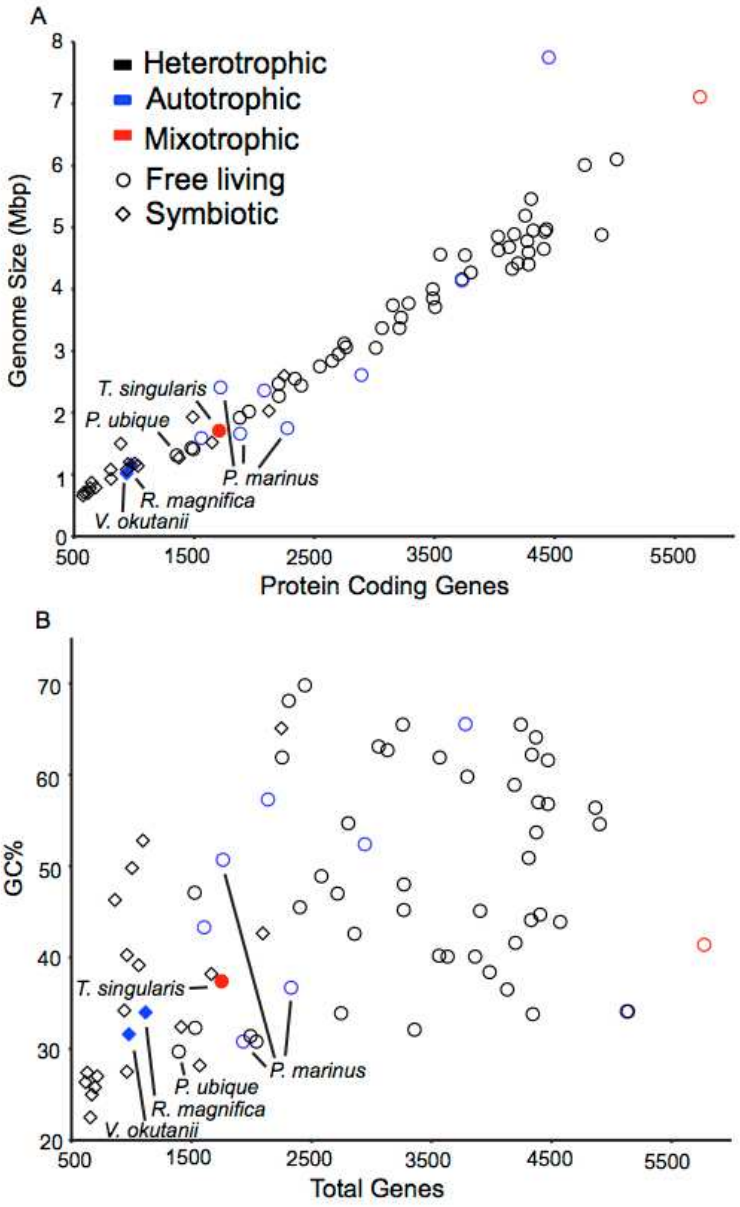

Figure 2.S2. Streamlined *T. singularis* genome. A) Number of protein coding genes versus genome size. B) Number of genes versus GC% [1].

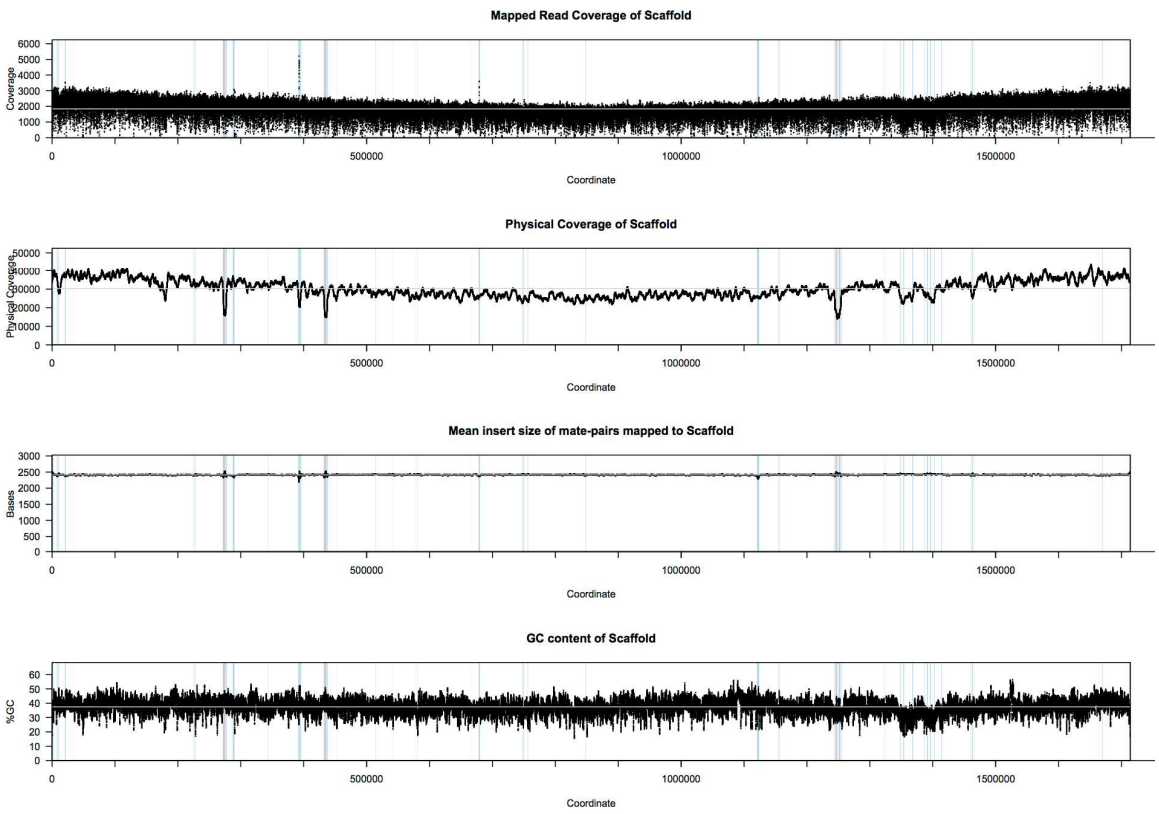

Figure 2.S3. *T. singularis* GSO-PS1 genome assembly plots. Read coverage, physical coverage, insert sizes of mate pairs, and GC% covering the *T. singularis* GSO-PS1 genome (top panel to bottom panel) [1].

Chapter 3. CO-OCCURING AUTOTROPHIC AND MIXOTROPHIC SUP05 SUBCLADES

Katharine T Marshall, Chris L Dupont, Clara A Fuchsman, Robert M Morris (In progress)

3.1 ABSTRACT

Members of the SUP05 clade of gamma-proteobacteria are ubiquitous in seawater. They are abundant in diverse coastal, pelagic, surface and deep ocean ecosystems and often dominate low oxygen waters, where they have key roles in the marine carbon, nitrogen, and sulfur cycles. Previous metagenomic studies have indicated that there are at least two lineages, originally identified by 16S rRNA sequences from symbionts of deep-sea clams and mussels and from environmental clones. We used the complete genome sequences of autotrophic and mixotrophic isolates to identify the genetic potential of co-occurring SUP05 subclades in diverse oxygen minimum zones (OMZs). We find that autotrophic SUP05 cells are facultative anaerobes and have the potential to fix carbon, oxidize sulfur, and reduce nitrogen to nitrate, nitrite, nitrous oxide, and dinitrogen gas. These metagenomic data from low oxygen regions provide the first evidence of complete denitrification in the SUP05 clade, and suggest that energy derived from sulfur-oxidation fuels autotrophic nitrogen loss from marine systems. Mixotrophic SUP05 cells are aerobic and have the potential for autotrophic and heterotrophic growth using organic and inorganic substrates. The enhanced phylogenetic and comparative genomic analyses we obtained using autotrophic and mixotrophic type strains (*T. autotrophica* SUP05-EF1 and *T. singularis* GSO-PS1) suggest that the SUP05 clade is comprised of co-occurring autotrophic and mixotrophic subclades.

3.2 INTRODUCTION

Oxygen minimum zones (OMZs) are unique biological and chemical regions that are expanding in the world's oceans (1, 2). In OMZs, aerobes with relatively high energy yielding metabolisms are replaced by anaerobic and microaerophilic microbes that rely on less energetically and biosynthetically efficient metabolisms. In these regions metabolisms tend to be much more diverse with oxidized forms of nitrogen, sulfur, manganese, iron and carbon dioxide all used as terminal electron acceptors (3). Many microbes use these chemicals for the heterotrophic breakdown of organic carbon. Chemoautotrophs can also become abundant, some couple sulfur oxidation to nitrogen reduction to yield energy for autotrophy (3). The shifts in carbon, nitrogen and sulfur cycling make OMZs fascinating regions for understanding unique metabolic strategies and nutrient cycling.

In the absence of oxygen, nitrate is the most effective electron acceptor and yields a similar amount of free energy as oxygen (3) making nitrogen reduction common in OMZs. Denitrification ($NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$) is the complete reduction of nitrate to dinitrogen gas and results in a loss of nitrogen by making it inaccessible to most organisms and allowing it to escape to the atmosphere. While OMZs make up ~1 percent of the total ocean volume, they represent 30-50 percent of nitrogen loss from the world's oceans (4–6). Nitrogen is a key component of many molecules necessary for life and the availability of nitrate limits primary production in many parts of the oxygenated ocean (7–9). The loss of nitrogen in any form represents a potential reduction in primary productivity in the ocean.

Nitrogen reduction does not always follow the traditional denitrification pathway. Bacteria may use partial denitrification, for example, only reducing nitrate to nitrite (3, 10, 11),

or may use one of the other known nitrogen reduction pathways: anammox (ANaerobic AMMonia OXidation) or DNRA (Dissimilatory Nitrate Reduction to Ammonia). Anammox ($NH_4^+ + NO_2^- \rightarrow N_2$) has only been found in planctomycetes, which have a specialized organelle, the anammoxosome, used to contain the toxic intermediate hydrazine made during the reaction (12, 13). DNRA ($NO_3^- \rightarrow NO_2^- \rightarrow NH_4^+$) has been found in more diverse groups of bacteria. DNRA and traditional denitrification share the first step ($NO_3^- \rightarrow NO_2^-$) (3) so unless a genome is complete or there are genes for other denitrification steps, the presence of the genes for this step (*nar* or *nap*) could mean that an organism is capable of partial denitrification, full denitrification or DNRA. Different nitrogen reduction pathways have different implications for the nitrogen cycle since only full denitrification and anammox result in the loss of dinitrogen gas.

Denitrification in OMZs is typically dominant as a heterotrophic process, with organic carbon being oxidized. Chemodenitrification by autotrophs, however, occurs when denitrification is coupled with the oxidation of a chemical, most frequently sulfur (3). When conditions become euxinic (characterized by the presence of sulfide), certain microbes couple denitrification with sulfide oxidation. Canfield et. al. broadened the geographic possibility for chemodenitrification by showing that genes for sulfur oxidation and reduction were present in OMZs with no measurable sulfide. This is fascinating because sulfate is reduced before the more favorable nitrate is depleted. The reduced sulfur is then immediately reoxidized, making a cryptic sulfur cycle that is coupled to denitrification (14).

Members of the SUP05 clade often dominate microbial communities in hypoxic waters (15-17). Growth experiments and whole genome analysis of the first autotrophic SUP05 islolate *T. autotrophica* strain SUP05-EF1 (Shah, in prep) indicate anaerobic reduction of nitrate to nitrite and of nitrous oxide to nitric oxide.Whole genome analysis of the first mixotrophic SUP05

isolate, *T. singularis* strain GSO-PS1 indicate that the potential for autotrophic and heterotrophic growth is ubiquitous (Chapter 2). Environmental metagenomes and single cell genomes support these findings (18–21). However, high sequence similarity within the SUP05 clade has limited attempts to resolve metabolic differences in co-occurring SUP05 subclades. Here we use the complete genome sequences of the type strains (*T. autotrophica* strain SUP05-EF1 and *T. singularis* strain GSO-PS1) to differentiate between co-occuring autotrophic and mixotrophic SUP05 subclades in nine OMZs. We report unique metabolic capabilities associated with autotrophic and mixotrophic lineages, including the potential for autotrophic denitrification and for oxygen driven niche partitioning.

3.3 Materials and Methods

Metagenomic reads from nine OMZ locations were pooled and assembled using Celera according to Dupont, 2012 (22) with the following modifications: Mer size was 14, error rate was increased to 92%, unitigger was set to BOG to allow for usage of Sanger and Illumina paired end and mate pairs. Contigs over 5000 bp were binned by 5-mer usage, a principle component analyisis (PCA) was run on them and the results were visualized. Contigs were aligned to reference genomes using Geneious (v. 7.1.7) and hits were used to identify *T. autotrophica* and *T. singularis* clusters. The 5-mer analysis, PCA and visualization were rerun this time including *T. singularis* GSO-PS1 and *T. autotrophica* SUP05-EF1 genomes (Appendix C) ORFs on contigs were predicted and annotated using the JCVI pipeline and RAST (v. 2.0, NMPDR). Each genome was blasted against the predicted genes in its respective cluster (e^{-10} cut off) to find genes present in the cluster but not in the genome. Comparison of key metabolic gene suites was done according to Chapter 2.

The synteny of key genes was compared using RAST and the identity of key genes were checked by blasting or making maximum likelihood trees using RAxML. Trees were made for *narG*, *nirK*, *norB*, *nosZ*, (Appendix C) and ribulose bisphosphate carboxylase (RuBisCo) (Figure 4). Contigs less than 5000 bp were searched for RuBisCo genes to add to the phylogenetic tree in order to capture more diversity of RuBisCos in the OMZ samples.

3.4 RESULTS

A metagenomic assembly was compiled from nine OMZs (Figure 1, Table 1). This resulted in 204,511 contigs totaling 454 Mbp of SUP05 environmental sequence data. Of a total of 13,285 contigs >5000 bp in the dataset, 195 were assigned to *T. singularis* GSO-PS1 (mean coverage 9X) and 350 were assigned to *T. autotrophica* SUP05-EF1 (mean coverage 15X). Only 78 (5.5%) of the 1,430 proteins assigned to the GSO-PS1 associated cluster, were not found in the GSO-PS1 genome (e value cut off at e^{-10}). Proteins identified in the GSO-PS1 associated OMZ cluster but not in the GSO-PS1 genome included, one phage related gene, one autotrophy-linked gene, and 37 hypothetical proteins. A much larger fraction (14.4%), of the 3,779 proteins assigned to the SUP05-EF1 associated cluster were not found in the SUP05-EF1 genome. Proteins identified in SUP05-EF1 associated OMZ cluster, but not in the SUP05-EF1 genome included 11 phage related genes, 12 cytochrome C oxidase related genes, 9 autotrophy-linked genes, 8 urease genes and 155 hypothetical proteins.

The metabolic gene suites found in OMZ clusters mirrored those found in the *T. singularis* and *T. autotrophica* genomes (Figure 2). Comparison of the *T. singularis* GSO-PS1 genome to OMZ sequences in the GSO-PS1 associated cluster indicate that this group has the conserved genetic potential for aerobic mixotrophy, including organic carbon metabolisms using sugar alcohols, organic acids, mono- and oligo-saccharides and fermentation, as well as the

potential for autotrophic carbon fixation and sulfur oxidation. Comparison of the *T. autotrophica* SUP05-EF1 genome to OMZ sequences in the SUP05-EF1 associated cluster indicate this group has the genetic potential for chemoautotrophic nitrogen reduction and enhanced sulfur oxidation. Furthermore, 4.5% of the total proteins found the GSO-PS1 associated cluster were ABC transporters compared to only 2.1% of the proteins found in the SUP05-EF1 associated cluster, suggesting that *T. singularis* have better ability to transport organic carbon into their cells for heterotrophic use.

Analyses of 31 sulfur oxidation genes indicate that the metabolic potential of OMZ SUP05 associated clusters have sulfur oxidation pathways that are similar to the type strains and correspond to differences in sulfur oxidation potential of previously defined subclades. The *T. autotrophica* SUP05-EF1 associated cluster contains *sox* (except *soxCD*), complete *dsr*, *aprAB, sat,* and *rhodanese* genes, which gives this group the potential to oxidize thiosulfate or elemental sulfur to sulfate. The *T. singularis* GSO-PS1 associated cluster contains *aprB*, a subset of *dsr* genes, *sat,* and *rhodanese,* which gives this group the potential to oxidize a smaller subset of reduced sulfur substrates, including thiosulfate to sulfite and an APS intermediate to sulfate. Though aprA was not detected using these methods, if it exists in this group, they would also have the ability to oxidize sulfite to APS, giving them potential for oxidation of thiosulfate to sulfate. All sulfur oxidation genes identified in the SUP05-EF1 and GSO-PS1 genomes were also identified in OMZ clusters.

SUP05 cells have Type IA, type II or both forms of RuBisCO (Figure 3, 4). The RuBisCo type II in the *T. autotrophica* SUP05-EF1 associated cluster were more diverse than has been found previously with one of the most diverged sequences on OMZ Contig F, which also contains the nitrous oxide reduction genes (*nos*). The *T. autotrophica* SUP05-EF1 isolate is the

first SUP05 found to have two RuBisCos. Its RuBisCo type II is similar to other autotrophic SUP05 RuBisCos. Its RuBisCo type IA is similar to that of mixotrophic SUP05. These Rubiscos are more specifically type IAq: each has a type IAq-specific six amino acid motif and each has cbbO and cbbQ near it (23). As is typical for organisms with RuBisCo IAq, *T. singularis* GSO-PS1 does not have carbon-concentrating mechanisms such as carbonic anhydrase. *T. autotrophica* SUP05-EF1 does have carbonic anhydrase, though it is likely associated with the RuBisCo II and not the RuBisCo IAq. There were no RuBisCo IAq found in the *T. autotrophica* SUP05-EF1 associated OMZ cluster and no RuBisCo II found in the *T. singularis* GSO-PS1 associated OMZ cluster.

Analysis of *T. autotrophica* genomes and associated OMZ contigs shows that members of the *T. autotrophica* clade have different nitrogen reduction capabilities (Figure 3, 5, 6). The complete genome of the *T. autotrophica* SUP05-EF1 has the genetic potential to reduce nitrate to nitrite (*nap* and *nar*) and nitric oxide to nitrous oxide (*nor*) and the SI (Saanich Inlet) metagenome to reduce nitrate to nitrous oxide (*nap*, *nar*, *nir*, and *nor*) (Figure 5). The *T. autotrophica* associated OMZ cluster contained genes for all of these steps plus reduction to N_2 gas (*nar*, *nap*, *nir*, *nor* and *nos*), which has not previously been found in SUP05. Genes for nitrogen reduction were found on 13 contigs, with varying numbers of the key genes in each step found: 3 narGH, 3 napAB, 4 nirK, 3 norCB, and 1 nosZ. One nirK was 21% of the length of the other nirK in the OMZ contigs. Most of the key nitrogen reduction genes were found on contigs with other key or peripheral nitrogen reduction genes.

The synteny of nitrogen reduction suites in *T. autotrophica* is diverse further indicating that *T. autotrophica* have different nitrogen reduction capabilities. Evidence of this is found next to the well-defined *T. autotrophica* RuBisCo genes, adding certainty that, though the nitrogen

reduction genes are variable, they belong to *T. autotrophica* (Figure 3). RuBisCo genes for autotrophy are found on two contigs in the OMZ cluster associated with *T. autotrophica*, near nitrate (*nar* or *nap)* and/or nitrous oxide (*nos)* reduction genes. OMZ contig F contains nitrous oxide reduction genes (*nosZDFLY)* on a contig with nitrate reduction genes (*napGH*) and RuBisCo. The reads forming this contig come from Landsort Deep in the Baltic Sea and have 30X coverage. OMZ contig G has a nitrate reduction gene (*nar)* near the RuBisCo, similar to the other SUP05 genomes and metagenomes. In this region, OMZ contig G and *T. autotrophica* SUP05-EF1 both have nitrate/nitrite transporters and all sequences compared contain at least one molybdopeterin cofactor biosynthesis gene, further suggesting synteny of nitrogen reduction genes and RuBisCo between *T. autotrophica*.

The key gene for the reduction of nitrite to nitric oxide (*nirK)* was found on four *T. autotrophica* associated OMZ contigs (contigs A-D) (Figure 6). The synteny and length of the *nirK* were variable. OMZ contig C and SI have short *nirK* genes with 80/82 AAs respectively rather than ≥378 AAs in others. The shortened *nirK* genes contain one multicopper oxidase pfam instead of two like the longer *nirK* genes. *T. autotrophica* SUP05-EF1 has nitric oxide reduction genes (*nor)* but completely lacks nitrite reduction genes (*nir)*. The presence of copper metallochaperone and/or prion proteins in all sequences except OMZ contig D suggests that these nitrite and nitric oxide reduction genes are frequently syntenous between *T. autotrophica*.

3.5 DISCUSSION

Genetic analysis of SUP05 from OMZs confirms and expands our knowledge of their metabolic diversity. Our data suggest that *T. autotrophica* is more diverse than *T. singularis* because a higher percentage of genes found in the *T. autotrophica* associated OMZ cluster are missing from the *T. autotrophica* SUP05-EF1 genome (14.4%) versus the *T. singularis*

associated OMZ cluster from the *T. singularis* GSO-PS1 genome (5.5%). This greater diversity is supported by more diverse 16S sequences than *T. singularis* (Intro, Figure 1) and the differing lifestyles of *T. autotrophica* as symbionts and free living (17, 24–26). *T. singularis* genomes from the Puget Sound and North Atlantic surface waters were at least 92.2% identical at the amino acid level but only 72.1% identical to genomes from the South Atlantic deep waters (Chapter 2). While this suggests there may be a surface and deep ecotypes in the *T. singularis* subclade, we could not detect them using the OMZ dataset.

The *T. autotrophica* and *T. singularis* associated OMZ clusters have similar metabolic pathways to their respective representative genomes, showing that these genomes are suitable representatives of the subclades (Figure 2). For *T. autotrophica*, these data corroborate with metagenomes collected from sites where this subclade is abundant, suggesting that they are chemoautotrophs that can couple sulfur oxidation to nitrogen reduction in the absence of oxygen (18, 20). For *T. singularis*, these data corroborate with isolate growth experiments and single cell genomes that suggest they are aerobic sulfur oxidizing chemomixotrophs (19, 21, 27). In previous metagenomic analyses from locations where these subclades co-occur, they have been lumped together because they are similar at the amino acid level and hard to separate (28). In this study, we clearly separated the subclades with a combination approach using genomes from cultured organisms and environmental data allowing us separate data from the subclades even when they co-occur.

Genes in the *T. autotrophica* associated OMZ cluster but missing from the *T. autotrophica* SUP05-EF1 genome suggest diversity of autotrophy, respiration (cytochrome C oxidase for oxygen) and possible horizontal gene transfer via phage. Growth experiments have shown that though *T. autotrophica* is a nitrogen reducer, it requires ammonia from the

environment (Shah et al, in progress). The presence of urease on a *T. autotrophica* associated OMZ contig suggests that some *T. autotrophica* can take up urea to produce ammonia (29).

Autotrophy using RuBisCo is common in proteobacteria, with over half having two or more copies of RuBisCo (23, 30). RuBisCo is not ideally suited to oxygenated environments because CO_2 and O_2 molecules are difficult for it to differentiate. RuBisCo has evolved forms that are specialized for high or low oxygen conditions (23). Having multiple RuBisCo genes allows bacteria to adapt to variable oxygen conditions, while bacteria that experience more stable oxygen conditions tend to have only one RuBisCo. Tcherkez et. el. hypothesize that RuBisCos are differentiated by how tightly the CO_2 binds to RuBisCo. If CO_2 binds tightly it more closely resembles a carboxylate group, which differentiates CO_2 from O_2. However, it can become so tightly bound that the reaction slows. Therefore, the catalytic rate and ability to distinguish CO_2 from O_2 are enmeshed and either RuBisCo is more non-specific (better for low oxygen) or it has slower kinetics (better for high oxygen) (31). Members of the *T. singularis* clade have RuBisCo IA, which is adapted to higher oxygen, while members of the *T. autotrophica* clade tend to have RuBisCo II which is adapted for lower oxygen (23) (Figure 3, 4). SUP05-EF1 is the first SUP05 shown to have both RuBisCo IA and II, suggesting that it is specialized for more variable or intermediate oxygen conditions where both RuBisCos are useful. The RuBisCo II (low oxygen) found on OMZ contigs F and G suggest that they may be more diverged groups of *T. autotrophica*.

While some bacteria do complete denitrification, others do partial denitrification and pass off their reduced product to other bacteria that can do the next step in denitrification, thus being part of a denitrifying community (3). So far, the only *T. autotrophica* data available that represents an individual comes from *T. autotrophica* SUP05-EF1, which performs two,

nonconsecutive, steps in denitrification (Shah et al, in prep) (Figure 5). In the environment, this isolate might be undergoing gene loss, using only its nitrate reduction gene and slowly loosing the presence and/or function of all other nitrogen reduction genes, or it might use nitrate and low levels of nitric oxide that leak from other nitrogen reducing bacteria. The SI metagenome shows that SUP05 are capable of more denitrification steps than are found in *T. autotrophica* SUP05-EF1 with its ability to reduce nitrite. In this study we show that *T. autotrophica* as a population are capable of all of the nitrogen reduction steps that have previously been found as well as the reduction of nitrous oxide to N_2 gas using *nos* genes. We have yet to determine if some *T. autotrophica* individuals are capable of complete denitrification or if they all perform only certain steps in the denitrification process thus needing to be a part of a denitrification community for nitrogen loss to occur (Figure 2).

The presence and synteny of denitrification genes is variable in the *T. autotrophica* clade (Figure 3, 6). The absence of genes on an OMZ contig could represent genes missing from a genome or in another section of a genome, thus analysis must be based on the presence of genes. The presence of nitrous oxide reduction genes, *nosZ* and peripheral *nos*, in OMZ contig F shows that some *T. autotrophica* have the ability to produce dinitrogen gas, representing potential nitrogen loss and impacting nitrogen cycling (Figure 3). The proximity of these denitrification genes to RuBisCo, which has well defined phylogeny in SUP05, provides high confidence that these genes actually belong to *T. autotrophica*. Determining if this ability is unique to Baltic populations and quantifying the portion of *T. autotrophica* with *nos* genes would help with evaluation of the environmental impact of *T. autotrophica*. *T. autotrophica* are abundant in the Baltic, reaching 30% of the total prokaryotes at the oxic-anoxic interface (32). Though *nap* genes are only shown on OMZ contig F, where present, the *nap* genes are very near the *nar* genes, ~14

kbp away in SUP05-EF1 and ~22 kbp away in the SI metagenome (~1-2% of the genome away). The molybdopterin cofactor biosynthesis gene near every *nar* and *nap* complex provides evidence of synteny, and are present because both *nar* and *nap* require a bis-molybdopterin guanine dinucleotide cofactor at the active site in order to function (33).

The key nitrite reduction gene (*nirK)* is present in the SI metagenome though it is on the edge of a contig and short, suggesting either that it is short or was cut off due to lack of sequence (Figure 6). Finding a similarly short *nirK* with similar syteny in the middle of OMZ contig C, suggests that the *nirK* on the SI metagenome contig is indeed short. Studies on other biological systems suggest that the short *nirK* genes may not function since multicopper oxidases normally use at least two domains and these have only one domain (34, 35). Further genetic analysis, new isolates and more growth experiments using *T. autotrophica* well help define the full diversity of nitrogen reduction in this subclade.

3.6 CONCLUSIONS

The SUP05 clade is a diverse group of organisms with a variety of metabolisms. This diversity may have major impacts on nutrient cycling. While no new essential metabolic genes were found for the *T. singularis* associated cluster in the OMZs, genes show that globally *T. singularis* is a unique SUP05 subclade that is obligately aerobic, mixotrophic, with an incomplete sulfur oxidation pathway and phototrophic potential. The *T. autotrophica* associated cluster contained genes providing the first evidence of complete denitrification. Differences in nitrogen reduction and RuBisCo genes in *T. autotrophica* suggest that this subclade may have more metabolically distinct ecotypes within it that are specialized for different oxygen conditions. SUP05-EF1 has both high and low oxygen RuBisCo and genes for nitrate and nitric

oxide reduction steps, suggesting that it would be a strong competitor in variable or intermediate oxygen conditions such as an oxycline.

T. autotrophica RuBisCos and nitrogen reduction genes indicate that many *T. autotrophica* are specialized for lower oxygen conditions. RuBisCo II (low oxygen) has been found in *T. autotrophica* in many environmental studies without evidence of RuBisCo IA (high oxygen) and synteny supports that high oxygen RuBisCo is missing from most *T. autotrophica*. Nitrite and nitrous oxide reduction genes (*nir* and *nos)* in the *T. autotrophica* associated OMZ cluster show that some *T. autotrophica* are either capable of complete denitrification or are capable of different nitrogen reduction steps than *T. autotrophica* SUP05-EF1. A complete denitrification pathway would make them better competitors in low oxygen environments. Isolates and complete genomes of *T. autotrophica* with different RuBisCo and denitrification genes suites are needed, particularly *T. autotrophica* with the ability to produce N_2 gas to learn about their role in nitrogen loss. Understanding the full genetic potential and doing comparative growth experiments with other isolates would allow us to better understand the diversity and environmental impact of this diverse clade.

3.7 Additional Information

Additional information for this chapter can be found in appendix C.

1. Stramma L, Johnson GC, Sprintall J, Mohrholz V (2008) Expanding Oxygen-Minimum Zones in the Tropical Oceans. *Science* 320(5876):655–658.

2. Diaz RJ, Rosenberg R (2008) Spreading Dead Zones and Consequences for Marine Ecosystems. *Science* 321(5891):926–929.

3. Lam P, Kuypers MMM (2011) Microbial Nitrogen Cycling Processes in Oxygen Minimum Zones. *Annu Rev Mar Sci* 3(1):317–345.

4. Codispoti LA, et al. (2001) The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? *Sci Mar* 65(S2). Available at: http://scientiamarina.revistas.csic.es/index.php/scientiamarina/article/viewArticle/684.

5. Gruber N (2004) The Dynamics of the Marine Nitrogen Cycle and its Influence on Atmospheric CO2 Variations. *The Ocean Carbon Cycle and Climate*, NATO Science Series., eds Follows M, Oguz T (Springer Netherlands), pp 97–148. Available at: http://link.springer.com/chapter/10.1007/978-1-4020-2087-2_4 [Accessed May 27, 2015].

6. Gruber N (2008) The Marine Nitrogen Cycle: Overview and Challenges. 1–50.

7. Falkowski PG (1997) Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean. *Nature* 387(6630):272–275.

8. Codispoti LA (1995) Is the ocean losing nitrate? *Nature* 376(6543):724–724.

9. Tyrrell T (1999) The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* 400(6744):525–531.

10. Gonzalez JM, Kiene RP, Moran MA (1999) Transformation of Sulfur Compounds by an Abundant Lineage of Marine Bacteria in the alpha -Subclass of the Class Proteobacteria. *Appl Env Microbiol* 65(9):3810–3819.

11. Zumft WG (1997) Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev* 61(4):533–616.

12. Strous M, et al. (1999) Missing lithotroph identified as new planctomycete. *Nature* 400(6743):446–449.

13. Niftrik LA van, et al. (2004) The anammoxosome: an intracytoplasmic compartment in anammox bacteria. *FEMS Microbiol Lett* 233(1):7–13.

14. Canfield DE, et al. (2010) A Cryptic Sulfur Cycle in Oxygen-Minimum–Zone Waters off the Chilean Coast. *Science* 330(6009):1375 –1378.

15. Zaikova E, et al. (2010) Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* 12(1):172–191.

16. Walsh DA, et al. (2009) Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. *Science* 326(5952):578–582.

17. Swan BK, et al. (2011) Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* 333(6047):1296 –1300.

18. Lavik G, et al. (2009) Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* 457(7229):581–584.

19. Sunamura M, Higashi Y, Miyako C, Ishibashi J, Maruyama A (2004) Two Bacteria Phylotypes Are Predominant in the Suiyo Seamount Hydrothermal Plume. *Appl Env Microbiol* 70(2):1190–1198.

20. Anantharaman K, Breier JA, Sheik CS, Dick GJ (2012) Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proc Natl Acad Sci* 110(1):330–335.

21. Swan BK, et al. (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci*. Available at: http://www.pnas.org/content/early/2013/06/25/1304246110 [Accessed July 18, 2013].

22. Dupont CL, et al. (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* 6(6):1186–1199.

23. Badger MR, Bek EJ (2008) Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *J Exp Bot* 59(7):1525–1541.

24. Peek AS, Feldman RA, Lutz RA, Vrijenhoek RC (1998) Cospeciation of chemoautotrophic bacteria and deep sea clams. *Proc Natl Acad Sci U S A* 95(17):9962–9966.

25. Newton ILG, et al. (2007) The Calyptogena magnifica Chemoautotrophic Symbiont Genome. *Science* 315(5814):998–1000.

26. Kuwahara H, et al. (2007) Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, Calyptogena okutanii. *Curr Biol CB* 17(10):881–886.

27. Marshall KT, Morris RM (2013) Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J* 7(2):452–455.

28. Murillo AA, Ramírez-Flandes S, DeLong EF, Ulloa O (2014) Enhanced metabolic versatility of planktonic sulfur-oxidizing γ-proteobacteria in an oxygen-deficient coastal ecosystem. *Aquat Microbiol* 1:18.

29. Price NM, Harrison PJ (1987) Comparison of methods for the analysis of dissolved urea in seawater. *Mar Biol* 94(2):307–317.

30. Beller HR, et al. (2006) Whole-Genome Transcriptional Analysis of Chemolithoautotrophic Thiosulfate Oxidation by Thiobacillus denitrificans under Aerobic versus Denitrifying Conditions. *J Bacteriol* 188(19):7005–7015.

31. Tcherkez GGB, Farquhar GD, Andrews TJ (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci* 103(19):7246–7251.

32. Glaubitz S, Kießlich K, Meeske C, Labrenz M, Jürgens K (2013) SUP05 dominates the gammaproteobacterial sulfur oxidizer assemblages in pelagic redoxclines of the central Baltic and Black Seas. *Appl Environ Microbiol*. Available at: http://aem.asm.org/content/early/2013/02/11/AEM.03777-12 [Accessed May 20, 2013].

33. Richardson* DJ, Berks BC, Russell DA, Spiro S, Taylor CJ (2001) Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. *Cell Mol Life Sci* 58(2):165–178.

34. Messerschmidt A, Huber R (1990) The blue oxidases, ascorbate oxidase, laccase and ceruloplasmin Modelling and structural relationships. *Eur J Biochem* 187(2):341–352.

35. Permyakov E (2009) *Metalloproteomics* (John Wiley & Sons).

36. Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*:btu033.

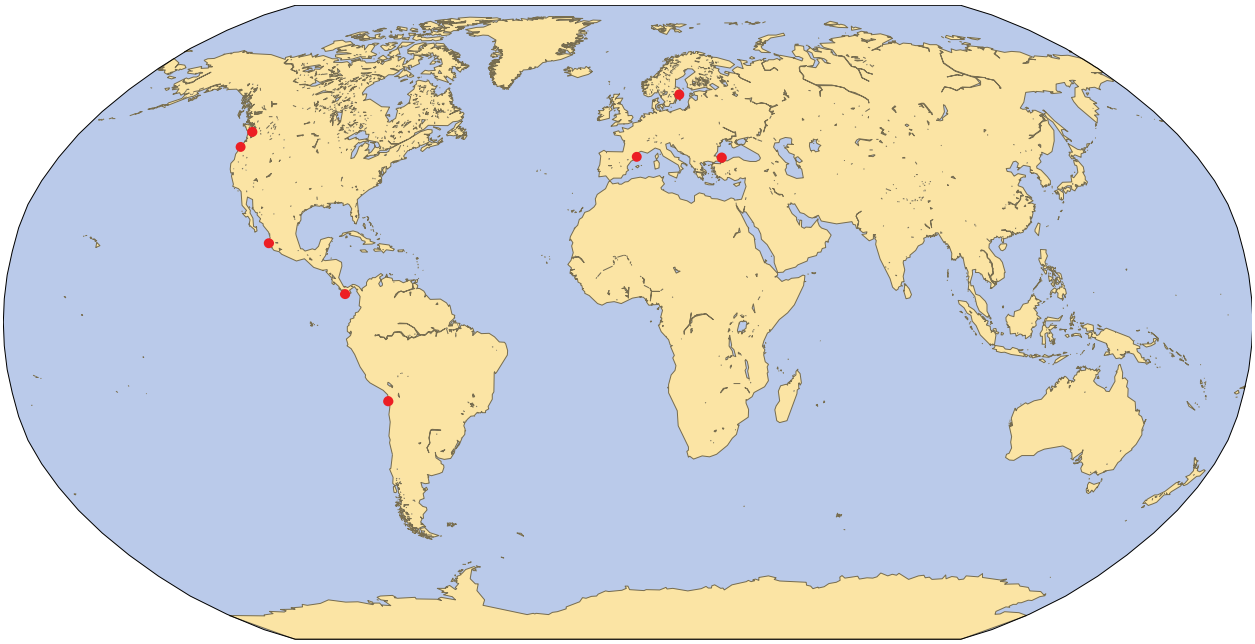

Figure 3.1. Locations of OMZ sample sites [1].

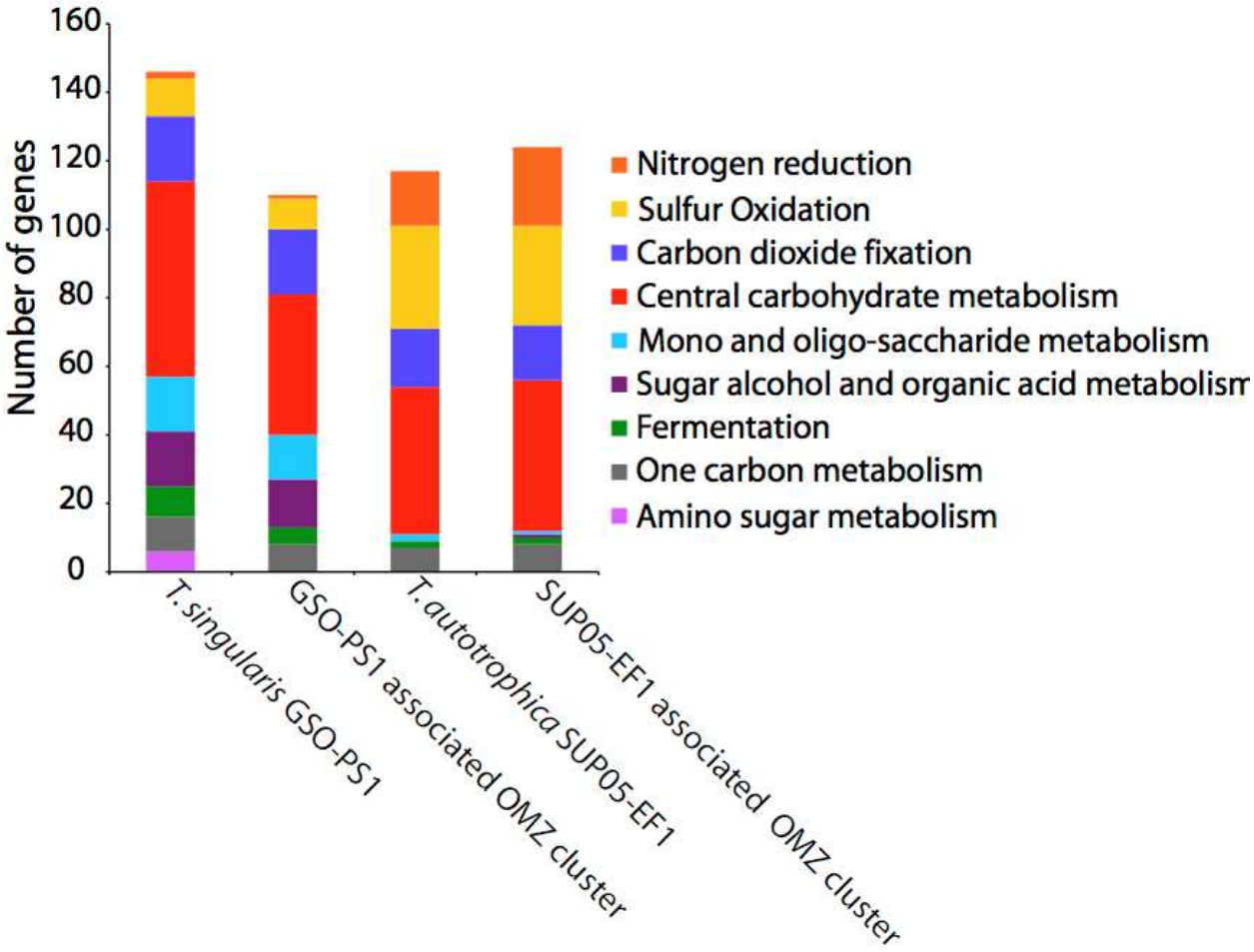

Figure 3.2. Comparison of genes found in major metabolic pathways in isolates and the OMZ clusters of *T. autotrophica* and *T. singularis*. Pathways include denitrification (nitrogen reduction), sulfur oxidation, autotrophy (carbon dioxide fixation), and catabolism (heterotrophy) (all other categories) [1].

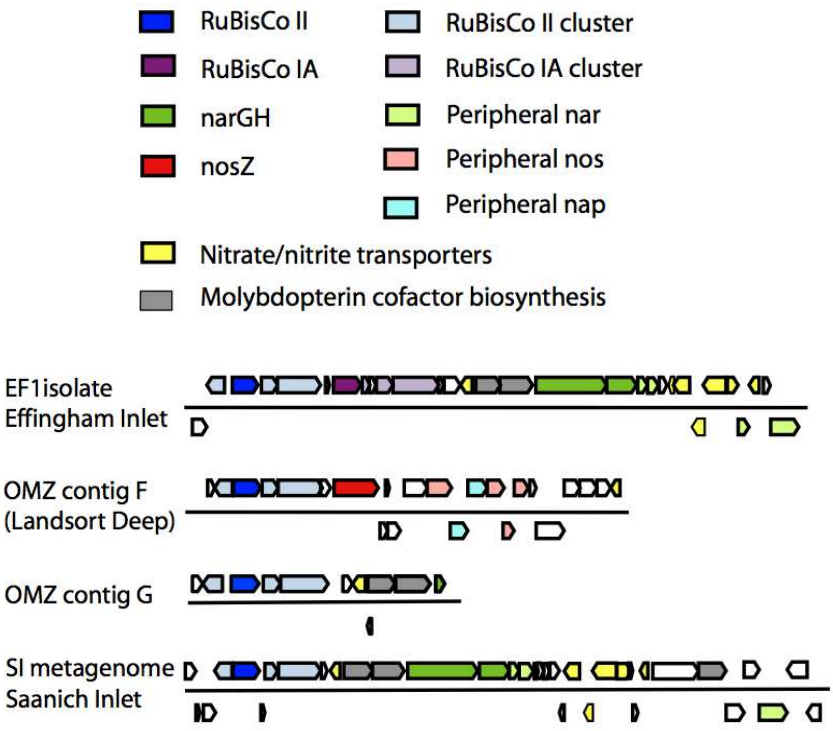

Figure 3.3. Synteny of RuBisCo genes found in *T. autotrophica* from OMZ contigs, SI metagenome and *T. autotrophica* SUP05-EF1 genome [1].

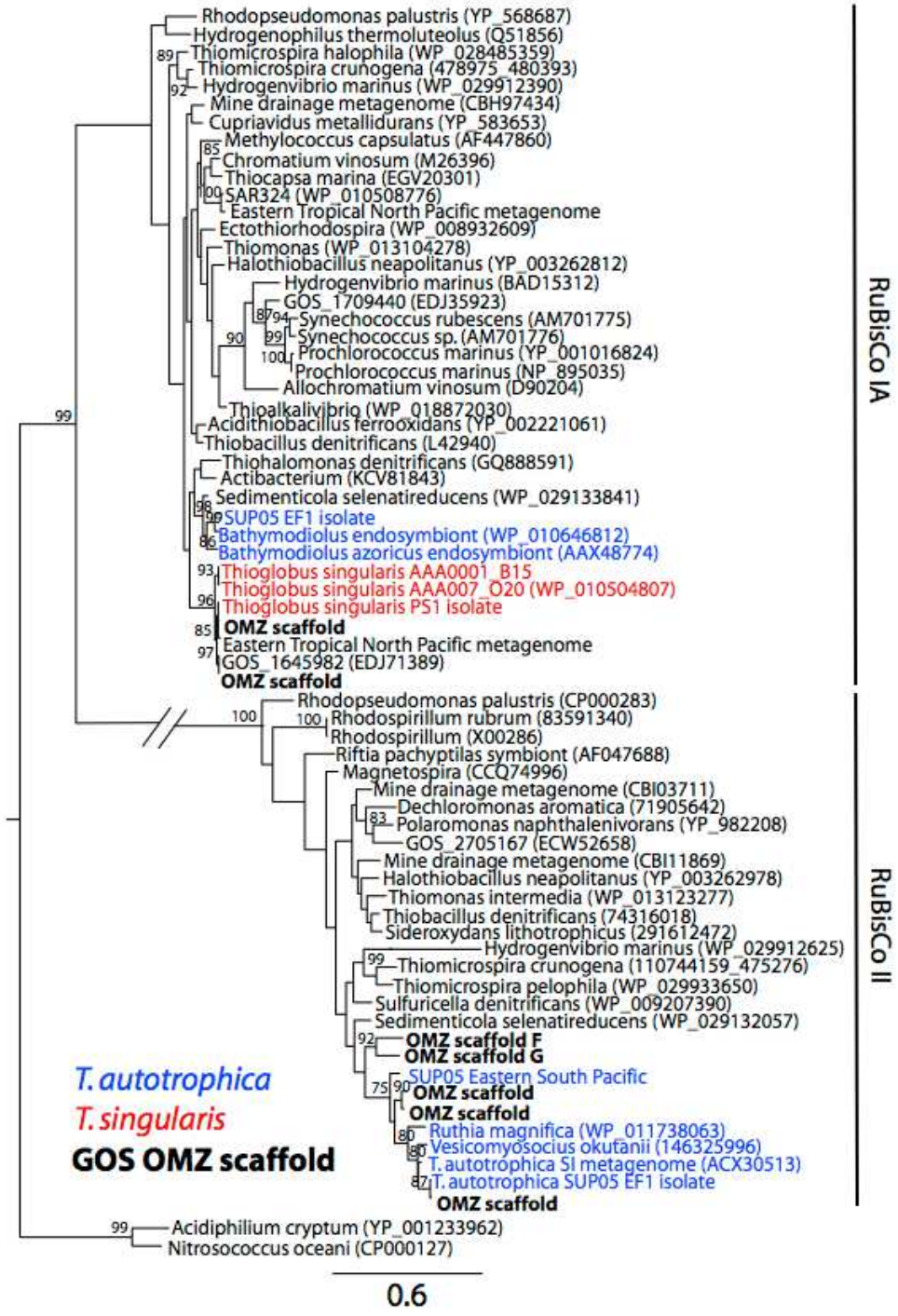

Figure 3.4. Phylogenetic tree of RuBisCo IA and II made using maximum likelihood in RAxML (36) with 100 bootstraps (≥75 shown) [1].

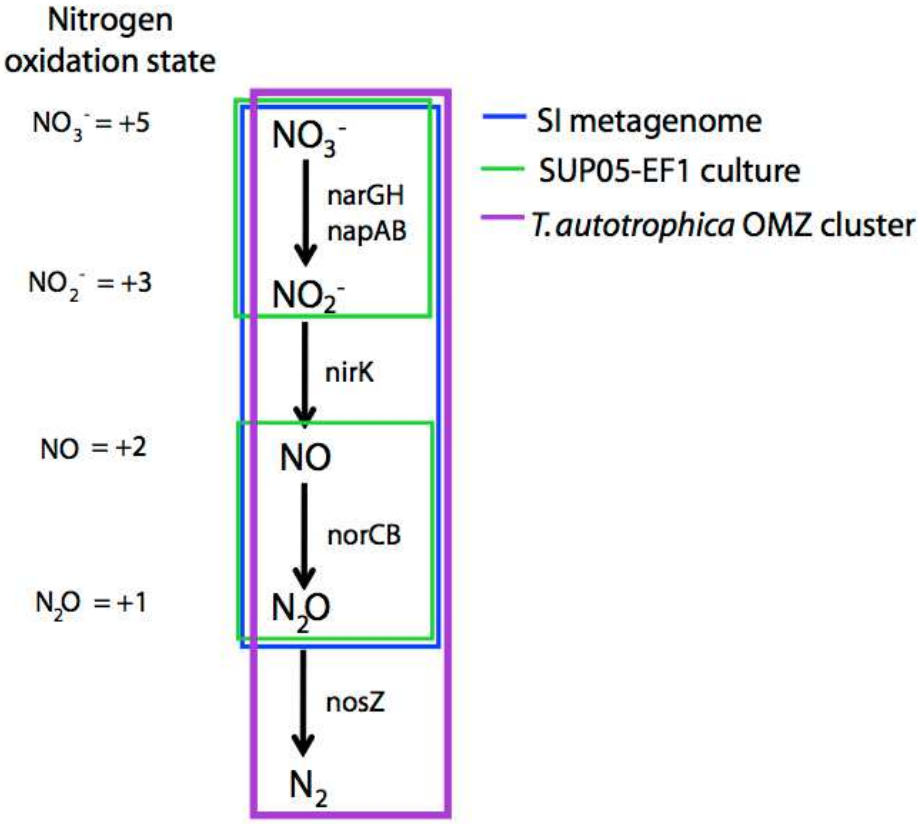

Figure 3.5. Comparison of genes associated with respiration by denitrification found in *T. autotrophica*: *T. autotrophica* SUP05-EF1 genome, SI metagenome, and OMZ cluster [1].

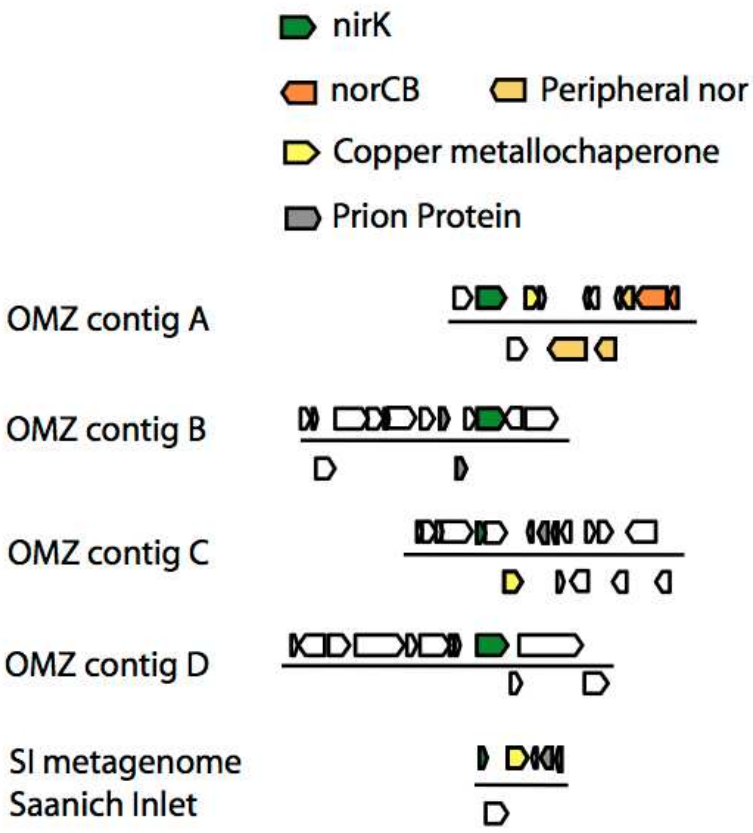

Figure 3.6. Synteny of nirK genes found in *T. autotrophica* from OMZ contigs and SI metagenome [1].

Table 3.1. Sample site information.

Name	Type of OMZ	Gbp of seq.	Sequencing Platform(s)	Depths Sequenced	Publications
Pacific Coast of Mexico	Coastal, suboxic	0.75	454	No data	
Pacific Coast of Panama	Coastal, hypoxic	0.75	454	50 m, oxycline	
Baltic Sea-Landsort Deep	Sulfidic basin	10	454, Illumina	78 m, base of oxycline	Dupont et al. in review
Black Sea	Sulfidic basin	0.75	454	90m, OMZ core	
Lake Ciso	Sulfidic basin	0.75	454	Oxycline top and bottom	
Lake Banyoles	Sulfidic basin	15	454, Illumina	Oxycline top and bottom	
Pacific Coast of Oregon	Coastal, suboxic	0.75	454	OMZ core	Allen et al. 2013
Pacific Coast of Chile	Coastal, anoxic	1.92	454	Multiple depths	Canfield et al. 2010
Saanich inlet	Sulfidic basin	0.23	Sanger fosmids	No data	Walsh et al. 2009

CONCLUSION

Members of the SUP05 clade are metabolically diverse and subclades have different ecological niches, with *T. singularis* being mixotrophic and *T. autotrophica* being autotrophic. The *T. singularis* GSO-PS1 genome has autotrophic pathways and heterotrophic pathways/transporters for biosynthesis and sulfur oxidation and phototrophy for energy production. Mixotrophic *T. singularis* appears to have less diversity between individuals as evidenced by similar 16S sequences (introduction), similar key metabolic genes found in the *T. singularis* GSO-PS1 isolate and in the environmental genomes (chapter 2), and similar sequences found in OMZs versus the *T. singularis* GSO-PS1 genome (chapter 3). However, amino acid differences between deep and surface single cell genomes point to the possibility of multiple mixotrophic ecotypes (chapter 2). Meanwhile, autotrophic *T. autotrophica* have a great amount of metabolic diversity between individuals as evidenced by the presence of different nitrogen reduction genes in different individuals (chapter 3), the presence of different types and numbers of RuBisCos (chapter 3) and supported by greater 16S diversity (introduction) and lifestyle diversity.

These findings suggest that mixotrophic *T. singular*is are generalists that use multiple metabolic pathways to survive in varying conditions in different parts of the ocean. Maintaining multiple metabolic pathways is energetically expensive, but the genome is streamlined to reduce overall energy costs. Though this subclade may not be ideally suited for any one condition, its presence in varying environments and our metabolic data suggest that it could be competitive in diverse environmental conditions. Increasing numbers of bacterial clades have been found to use some form of mixotrophy so the *T. singularis* isolate may be an excellent model for this life

strategy. Autotrophic *T. autotrophica* have become more specialized for low to no oxygen conditions. The varying suites of nitrogen reduction genes and RuBisCo genes suggest that this subclade actually has multiple subclades within it that are specialized for different low oxygen conditions. The high abundance of autotrophic SUP05 throughout OMZs and absence in other pelagic environments supports this conclusion.

These findings have major implications for nutrient cycling. Sulfur oxidation is generally thought of as a process that occurs in or near anoxic/hypoxic environments, but mixotrophic SUP05 *(T. singularis)* is an example of an abundant group that oxidizes sulfur compounds in the oxygenated surface water. Autotrophic SUP05 (*T. autotrophica)* reduce nitrogen and they can be a nitrogen sink if they reduce nitrogen to dinitrogen gas. Both mixotrophic and autotrophic SUP05 are capable of primary production in the dark ocean, potentially adding significant organic carbon to the marine food web that is not accounted for by photosynthetic activity in the surface ocean. While ammonia oxidizing Archaea dominate primary production in the dark ocean, SUP05 have the potential to contribute significantly. The next steps in this research are to better define the ecological roles that are filled by mixotrophic and autotrophic SUP05. For mixotrophic SUP05 (*T. singularis)* this involves figuring out if metabolic genes are constitutively expressed suggesting they use all pathways all the time or if they switch which pathways they use depending on environmental conditions. For autotrophic SUP05 (*T. autotrophica)*, sequences of more genomes will help determine if all autotrophic SUP05 have partial denitrification pathways, or if some have complete denitrification pathways. Determining the abundance and distribution of autotrophic SUP05 with different metabolic potentials will help define the role of each subclade. The metabolic potential of mixotrophic and autotrophic SUP05 shows us that they have important roles in carbon, sulfur and nitrogen cycling in the

ocean, but the diversity of these pathways within the subclades makes further investigation necessary to understand their specific impacts on nutrient cycling and marine chemistry.

Appendix A

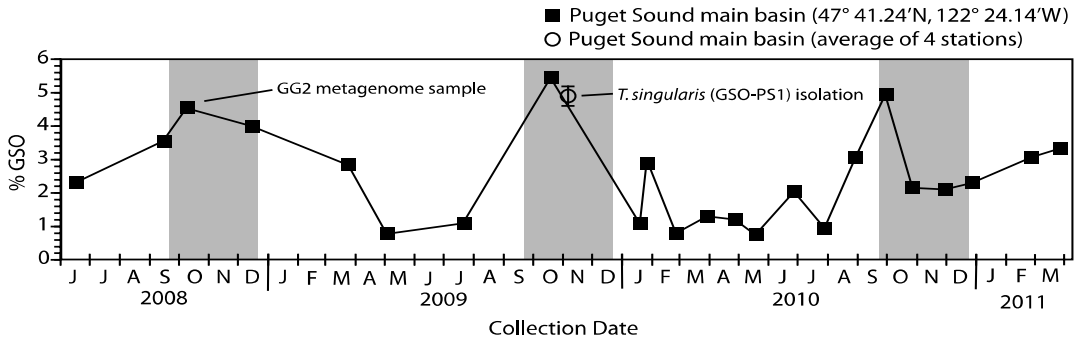

Figure AppA.1. Relative abundance of the Arctic96BD-19 subcluster in the Puget Sound main basin. Temporal shifts in abundance (filled squares). Average abundance from samples collected in November 2009 (open circle), when *T. singularis* GSO-PS1 was isolated. GSOs are most abundant in the fall (highlighted in grey) when *T. singularis* GSO-PS1 was isolated [1].

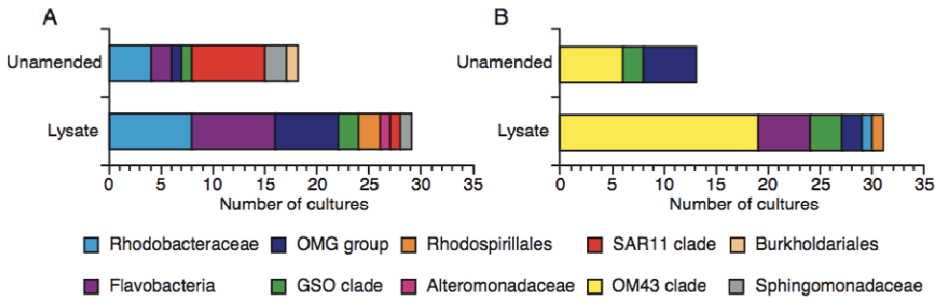

Figure AppA.2. High throughput isolation of marine bacteria. Number and identities of isolates obtained from a total of 216 inoculated wells per treatment. (A) From Puget Sound in unamended seawater media (control) and in seawater media amended with a mixed microbial cell lysate. (B) From the North Pacific gyre in unamended seawater media (control) and in seawater media amended with a diatom lysate [1].

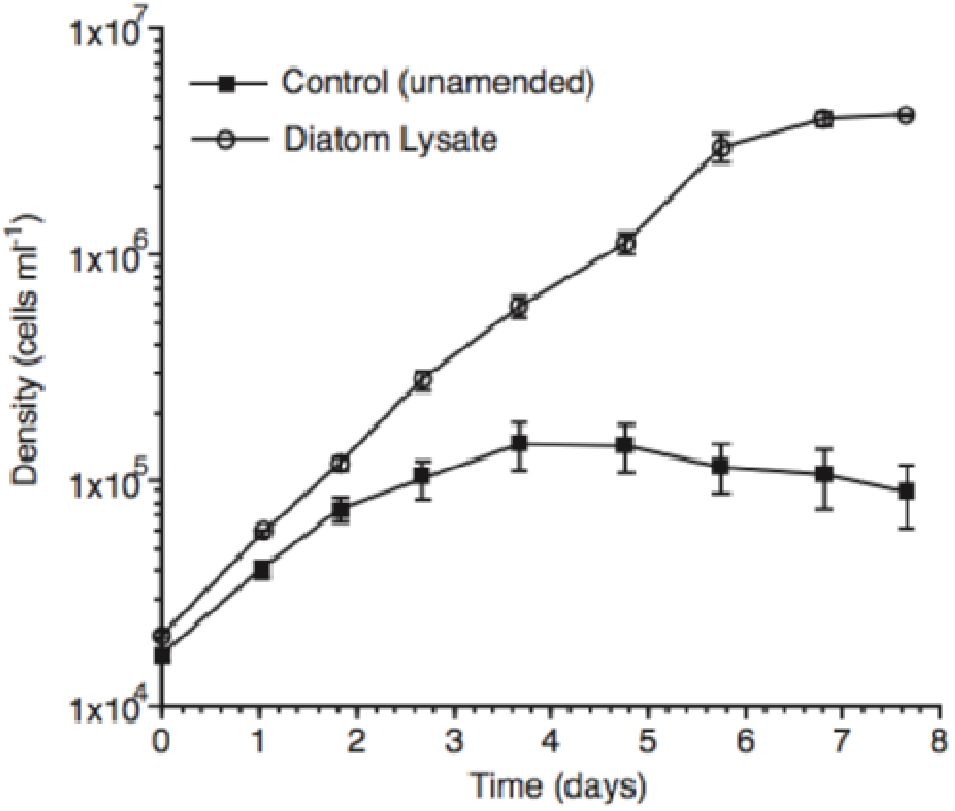

Figure AppA.3. Growth of *T. singularis* GSO-PS1 on filter sterilized seawater media (control), and media amended with diatom lysate. All experimental treatments were conducted in six replicate cultures and growth was plotted following initial detection (10^4 cells ml^{-1}) [1].

Appendix B

Table AppB.1. Genes that were not sequenced and assembled into the *T. singularis* GSO-PS1 genome from Solid sequencing but were completed and assembled into the genome using wet lab techniques.

Missing genes

Name	Start	Stop
tsf gene	10448	9551
fusA	272126	274244
carbohydrate kinase	386142	387079
membrane protein TolA	392559	393587
peptide chain release factor 2	420061	418967
putrescine/spermidine ABC transporter substrate-binding protein	491770	490676
GntR family transcriptional regulator: Disrupted gene	493336	494004
pseudo-pilin PulG	644557	644115
Hypothetical protein	688366	689580
ribosome hibernation promoting factor HPF	99847	700140
ABC transporter substrate binding protein	741327	740443
pilus assembly protein PilE	889455	888964
integrase	890502	889501
tRNA-Arg	890700	890610
DoxX subfamily	929864	929499
Hypothetical protein	930196	929933
metG gene: Disrupted gene	957046	955398
sugar ABC transporter substrate-binding protein	984417	983170
sugar ABC transporter substrate-binding protein	1019291	1018045
rplT	1140006	1139647
translation initiation factor IF-3	1140607	1140194
30S ribosomal protein S18	1163547	1163804
rplI	1163813	1164268
hypothetical protein	1239252	1238416
hypothetical protein	1251673	1248929
hypothetical protein	1254197	1251676
ABC transporter substrate-binding protein	1286670	1285657
sodium pump decarboxylase subunit gamma	1324588	1324355
FAD-linked oxidase	1356551	1355253
hypothetical protein	1452178	1452588

hisH	1453276	1452640
dehydrogenase	1466133	1467677
rpmE	1552436	1552224
BolA family transcriptional regulator	1552712	1552479
hypothetical protein	1570694	1570476
proteorhodopsin	1630487	1632194
hypothetical protein	1640152	1646991
tRNA-Thr	1680266	1680181
hypothetical protein	1681038	1680330
ubiquinol-cytochrome C reductase	1703839	1704471

Genes with holes

Name	**Start**	**Stop**
chromosomal replication initiator protein DnaA	1415	102
30S ribosomal protein S2	11533	10481
F0F1 ATP synthase subunit A	53830	54670
methylated-DNA--protein-cysteine methyltransferase	60560	60255
3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	61102	60557
rubrerythrin	170932	171351
spermidine/putrescine ABC transporter substrate-binding protein	180746	181820
ABC transporter ATP-binding protein	201308	202087
adenylylsulfate reductase subunit alpha	260087	258240
50S ribosomal protein L22	278790	279123
bifunctional proline dehydrogenase/pyrroline-5-carboxylate dehydrogenase	290584	294231
hypothetical protein	306869	307654
NADH dehydrogenase subunit D; disrupted	356173	357436
protein glxC	381279	381953
tRNA methyltransferase	391470	391505
excinuclease ABC subunit A	396259	399075
heme lyase subunit CcmF	407174	409093
DNA primase	421441	423267
peptidoglycan glycosyltransferase; disrupted	425127	423264
50S ribosomal protein L7/L12	439023	439403
DNA-directed RNA polymerase subunit beta	443613	447807
phosphopantothenoylcysteine decarboxylase; disrupted	454577	453411
RecX family transcriptional regulator	456660	457085
tgt	457113	458220
cystathionine gamma-synthase	470489	469266
fmt	473970	473028
cell division protein FtsY	477726	476773
membrane protein	477880	478319

amino acid ABC transporter substrate-binding protein	482699	481830
threonine synthase	484820	483747
ABC transporter ATP-binding protein	490607	489477
aminotransferase	494135	495522
glutamine amidotransferase; disrupted	497825	498540
BolA	506003	506305
peroxidase	542781	543384
acriflavin resistance protein	549075	552260
DNA-directed RNA polymerase subunit omega	586424	586768
phosphomethylpyrimidine synthase ThiC	591968	593836
acetoin utilization protein	599837	600778
oxidoreductase	619580	618576
hypothetical protein	619699	620753
aminopeptidase N	684354	686878
tRNA-dihydrouridine synthase A	687348	688884
polynucleotide phosphorylase/polyadenylase	726410	724312
RNA methyltransferase	773588	774346
biotin synthase	780414	781397
3-oxoacyl-ACP synthase	817730	816489
groEL	821889	820241
16S rRNA methyltransferase	833994	833230
ribonucleotide-diphosphate reductase subunit beta	841866	840723
ribonucleotide-diphosphate reductase subunit alpha	844209	841879
NAD+ synthetase	851363	853254
tRNA-Arg	877050	875204
NifS	928214	929449
pntB	949260	950693
2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1, 4-benzoquinol methylase	971307	970647
sugar ABC transporter ATPase	981306	980170
membrane protein	1038937	1039959
heme ABC transporter ATP-binding protein	1040106	1041676
peptide ABC transporter substrate-binding protein	1058268	1056642
cytidylate kinase	1061310	1060629
excinuclease ABC subunit C	1068368	1066575
hypothetical protein	1076477	1075871
2-oxoglutarate dehydrogenase E1	1115722	1118424
ammonium tranporter	1121225	1119822
hypothetical protein CDS	1154514	1152658
heme ABC transporter permease	1160992	1160252
30S ribosomal protein S6	1163148	1163537

iron ABC transporter ATP-binding protein; disrupted	1200822	1200107
hypothetical protein CDS	1204191	1203318
potassium transporter	1204393	1206352
trkA	1209956	1208567
dnaK	1214889	1216796
hypothetical protein	1226274	1224878
gatB	1227691	1226258
sodium:proton antiporter; disrupted	1237458	1236081
hypothetical protein	1258681	1257005
dihydrodipicolinate reductase	1305256	1304519
oxaloacetate decarboxylase	1324351	1322552
tRNA-Leu	1324748	1324655
hypothetical protein	1326143	1325634
hypothetical protein	1326414	1326088
sodium:calcium antiporter	1326512	1327462
hypothetical protein	1354520	1353396
membrane protein	1360858	1358879
hypothetical protein	1362764	1362426
glycosyl transferase	1363918	1362844
5'-methylthioadenosine phosphorylase	1387154	1386270
ADP-heptose--LPS heptosyltransferase	1393327	1392395
hypothetical protein; disrupted	1402855	1401023
xanthine permease	1426802	1428175
endonuclease IV	1451535	1452167
2-amino-4-hydroxy-6- hydroxymethyldihydropteridine pyrophosphokinase	1459781	1459279
3-dehydroquinate dehydratase	1502389	1501952
hypothetical protein	1524848	1522137
transcription-repair coupling factor	1542598	1546046
formate dehydrogenase subunit D	1591893	1591033
methylenetetrahydrofolate reductase	1599524	1598607
sugar ABC transporter substrate-binding protein	1612011	1610794
hypothetical protein	1639528	1640070
hypothetical protein	1716119	1715031

Appendix C

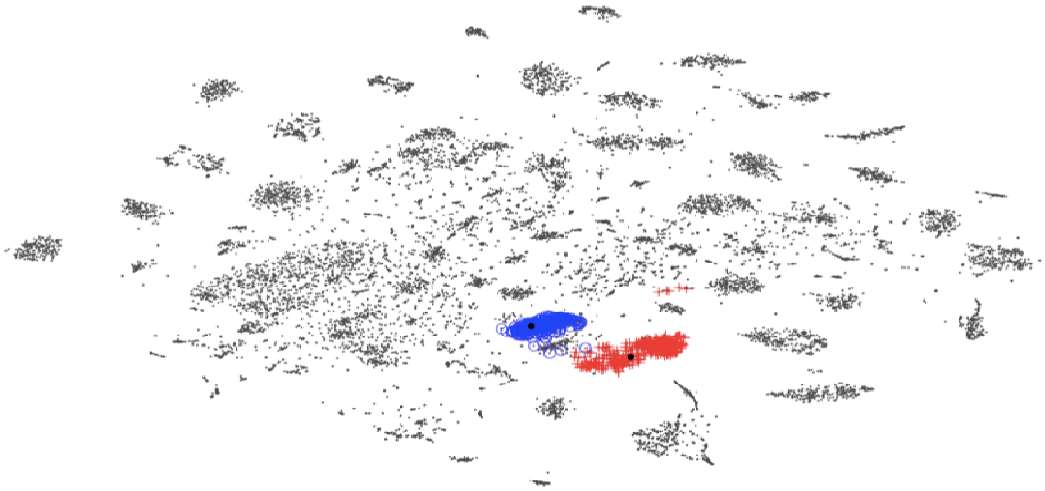

Figure AppC.1. Principle component analysis of contigs from nine OMZs. Contigs that align to *T. singularis* GSO-PS1 (Blue) and *T. autotrophica* SUP05-EF1 (red) are highlighted. Black dots represent the *T. singularis* GSO-PS1 and *T. autotrophica* SUP05-EF1 genomes in their respective clusters [1].

Figure AppC.2. Phylogenetic tree of NarG made using maximum likelihood in RAxML. OMZ contigs = boxed in green, SUP05-EF1 isolate = boxed in red (35) [1].

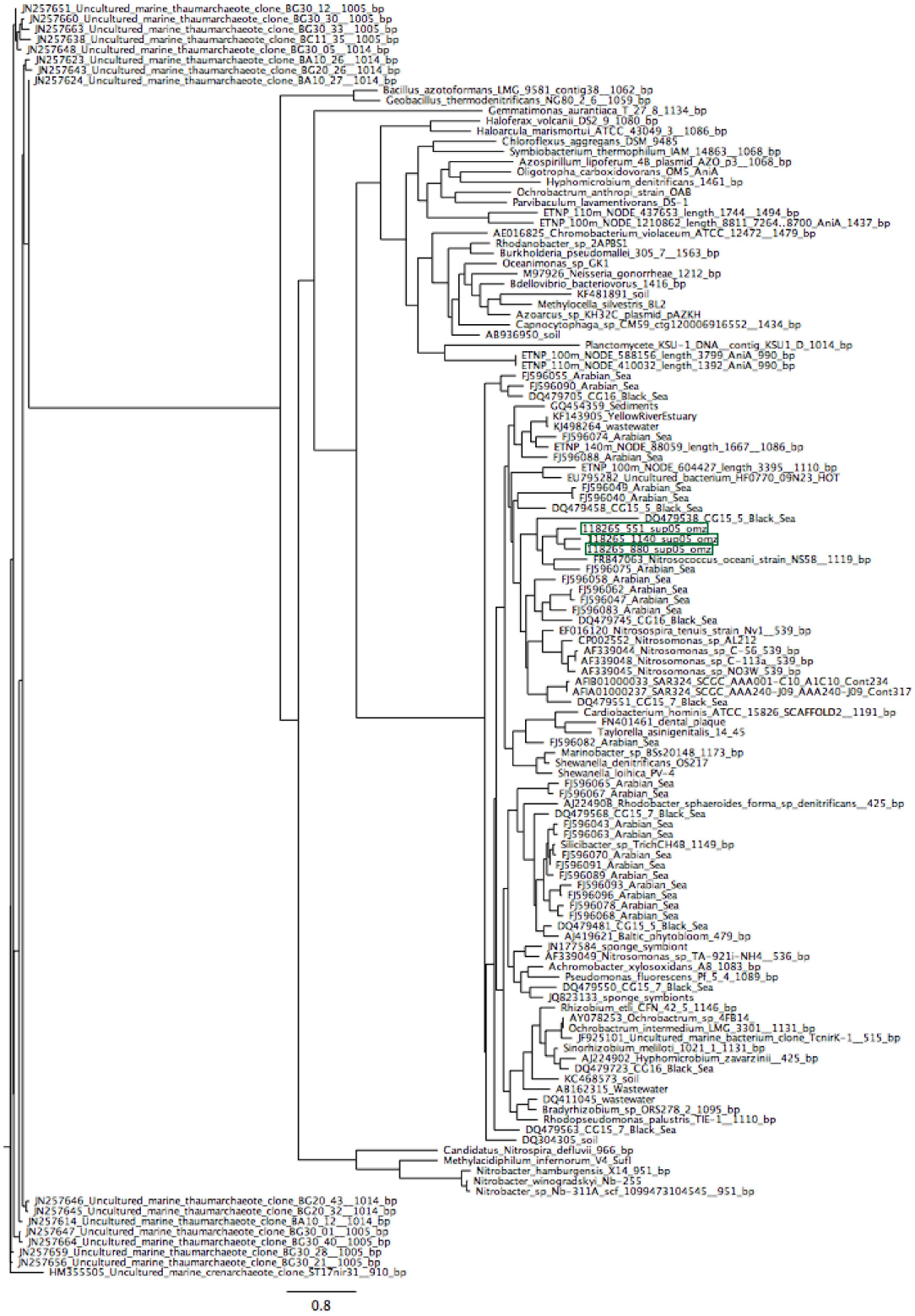

Figure AppC.3. Phylogenetic tree of NirK made using maximum likelihood in RAxML. OMZ contigs = boxed in green (35) [1].

Figure AppC.4. Phylogenetic tree of NorB made using maximum likelihood in RAxML. OMZ contigs = boxed in green, SUP05-EF1 isolate = boxed in red (35) [1].

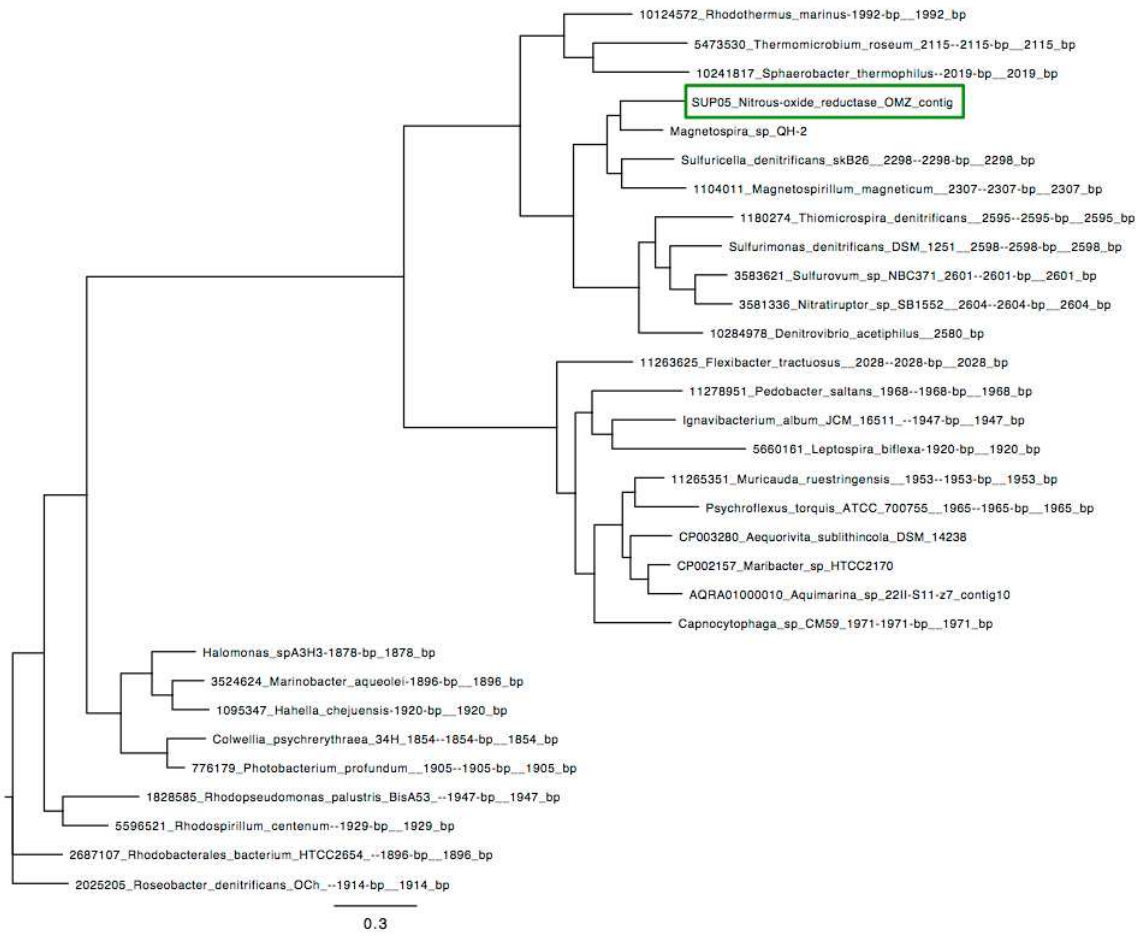

Figure AppC.5. Phylogenetic tree of NosZ made using maximum likelihood in RAxML. OMZ contig = boxed in green (35) [1].

VITA

KATHARINE MARSHALL LALISH
University of Washington
Seattle, WA 98105
kmarsha@u.washington.edu

PUBLICATIONS

Marshall KT, Morris RM (2013) Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J* 7(2):452–455.

Mattes TE, Nunn BL, **Marshall KT**, Proskurowski G, Kelley DS, Kawka OE, Goodlett DR, Hansell DA, Morris RM (2013) Sulfur oxidizers dominate carbon fixation at a biogeochemical hot spot in the dark ocean. *ISME J* 7(12):2349–2360.

EDUCATION

Anticipated Doctor of Philosophy in Biological Oceanography
University of Washington, Seattle, WA
Candidate for Doctor of Philosophy in Biological Oceanography
University of Washington, Seattle, WA (July, 2012)
Master of Science in Biological Oceanography
University of Washington, Seattle, WA (June 2011)
GPA: 3.74 (4.00 scale)
Bachelor of Arts in Biology
Occidental College, Los Angeles, CA (May 2008)
GPA: 3.82 (4.00 scale)/ 3.98 (4.00 scale)

Microbial Oceanography Course, Bermuda Institute of Ocean Sciences, Bermuda
♦ Participated in class, research, and presentations to learn methods and concepts in oceanography.

June-July 2008

Study Abroad Program, CIEE, Amsterdam, The Netherlands
♦ Selected to study globalization/environmental sustainability at the University of Amsterdam.

Aug 2006-Jan 2007

Study Abroad Program, Rotary Exchange, Manta, Ecuador
♦ Lived with an Ecuadorian host family and attended a local high school to learn about the culture and language.

Aug 2002-June 2003

RESEARCH EXPERIENCE

Research Cruise, University of Washington, North Pacific Ocean
♦ Collected bacteria and virus samples for graduate research.

May 2012

Research Cruise, University of Washington, Puget Sound
♦ Collected culturing samples for graduate research.

Nov 7-9, 2009

Research Cruise, Biological Institute of Ocean Sciences, Bermuda
♦ Collected culturing samples for graduate research.

Sept 1-4, 2008

Earth System Science Summer Fellow, University of California Irvine, Irvine, CA
♦ Wrote computer programs to do statistical analysis of small-scale physical variations in tropical tropospheric chemistry using the MOZAIC database. Attended lectures on Earth systems.

summer 2007

Summer Systematics Institute Intern, Academy of Sciences, San Francisco, CA
♦ Analyzed the genetics and morphology of tree frog specimens from the Gulf of Guinea Islands to determine speciation between populations.

summer 2006

Independent Plant Physiology Research, Occidental College, Los Angeles, CA
♦ Designed and carried out research on the physiological response to water stress of native and non-native species in Southern California.

spring 2007

Institute for Biodiversity and Ecosystem Dynamics Volunteer**,** Univ. of Amsterdam, NL
♦ Prepared soil samples for GC/MS analysis.

fall 2006

Independent Ecology Research, Occidental College, Los Angeles, CA
♦ Identified street trees and used GIS to map them for analysis.

spring 2006

Laboratory Technician's Assistant, Washington State University, Pullman, WA
♦ Digitized maps with GIS, helped with soil analysis, washed dishes.

summer 2003, 2004

Laboratory Technician's Assistant, University of Idaho, Moscow, ID
♦ Counted tree rings and prepared soil samples for stable isotope analysis.

summer 2003

PRESENTATIONS

Ocean Sciences Meeting, Honolulu Convention Center, HI
♦ Delivered an oral presentation of gamma sulfur oxidizer culturing research

Feb 2014

International Symposium on Microbial Ecology, Copenhagen Bella Center, Denmark
♦ Presented a poster of gamma sulfur oxidizer culturing research.

Aug 2012

Microbial Interactions in Marine Systems Symposium, Alfried Krupp College, Germany
♦ Presented a poster of gamma sulfur oxidizer culturing research.

July 2011

International Symposium on Microbial Ecology, Seattle Convention Center, WA
♦ Presented a poster of gamma sulfur oxidizer culturing research.

Aug 2010

National Conference on Undergraduate Research, Salisbury University, Salisbury, MD

♦ Presented a poster of tropospheric chemistry research from the previous summer.

April 2008

Earth System Science Poster Session, University of California Irvine, Irvine, CA
♦ Presented a poster of results from tropospheric chemistry analysis to the Earth System Science department.

summer 2007

National Conference on Undergraduate Research, Dominican University, San Rafael, CA
♦ Presented a poster of tree frog systematics research from the previous summer.

April 2007

Summer Systematics Institute Symposium, Academy of Sciences, San Francisco, CA
♦ Created a Power Point presentation of results from tree frog research and held a discussion with academy professors, staff and interns.

summer 2006

HONORS AWARDED FOR SCHOLARSHIP

GK-12 Teaching Fellowship, National Science Foundation

2012-2013

Graduate Research Fellowship, National Science Foundation

2009-2012

Early Career Researcher Poster Award, International Symposium for Microbial Ecology

Aug 2010

Phi Beta Kappa, Occidental College Chapter

2008

Dean's Scholarship, Occidental College

2004-2008

Dean's List, Occidental College

2005-2008

Mortar Board, Occidental College

May 2007-May 2008

Research Experience for Undergraduates funding, National Science Foundation

summers 2006, 2007

Chanda Morris Scholarship, Gritman Memorial Hospital

2004

TEACHING EXPERIENCE

Guest Lecture, South Seattle Community College, Seattle, WA
♦ Presented about my research to an introductory chemistry class

Nov 2014

Guest Lecture, Highline Community College MaST Center, Seattle, WA
♦ Presented about my research for a lecture series open to the public

Nov 2013

Guest Lecture, South Seattle Community College, Seattle, WA
♦ Presented about my research to an introductory chemistry class

Oct 2013

Guest Lecture, North Seattle Community College, Seattle, WA
♦ Presented about my research for a STEM lecture series open to students

April 2013

GK-12 Fellow, Roosevelt High School, Seattle, WA

♦ Taught and prepared lesson plans for 10th grade biology students.

Aug 2012- June 2013

Teacher's Assistant, University of Washington, Seattle, WA

♦ Taught two quiz sections for an oceanography course for nonscience majors.

Jan-March 2012

Undergraduate Research Mentor, University of Washington, Seattle, WA

♦ Trained undergraduates in my lab and worked with them on research.

fall 2010-Jun 2012

Substitute Instructor for Online Ocean 101 Course, University of Washington, Seattle, WA

♦ Answered student questions and graded exams and homework.

Aug-Sept 2011

Communicating Ocean Science Student Teacher, University of Washington, Seattle, WA

♦ Taught two 1st grade classes ocean sciences curriculum, designed one lesson.

spring 2011

Teacher's Assistant, University of Washington, Seattle, WA

♦ Taught two quiz sections for an oceanography course for science majors.

spring 2010

RE Volunteer Teacher, University Unitarian Universalists, Seattle, WA

♦ Taught 6th graders about different religions once a week.

Aug 2009-Jun 2010

Organic Chemistry Workshop Facilitator, Occidental College, Los Angeles, CA

♦ Wrote worksheets and taught/mentored organic chemistry students.

Jan 2007-May 2008

L.A. Bridges After-school Program Tutor, Occidental College, Los Angeles, CA

♦ Supervised and taught life-skills and art to as many as 20 middle school students.

fall 2004-spring 2006

Arts and Crafts Camp Counselor, Catalina Island Camps, Catalina Island, CA

♦ Designed art curriculum and provided for the well being of up to 12 girls.

summer 2005