PHYLOGENY AND PREDICTED FUNCTIONAL CAPABILITIES OF A SULFUR-OXIDIZING AND DENITRIFYING CLADE OF BACTEROIDETES FROM SULFIDIC ENVIRONMENTS

A Thesis

by

KECEN ZHOU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Jason B. Sylvan
Co-Chair of Committee,	Lisa Campbell
Committee Member,	Brendan E. Roark
Head of Department,	Shari Yvon-Lewis

August 2018

Major Subject: Oceanography

ABSTRACT

Environments rich in sulfur compounds (sulfidic) are common in the ocean, and the ability to gain energy (dissimilatory) from sulfur redox reactions is widespread in bacteria. The Sulfiphilic Bacteroidetes (SB), have been found exclusively in sulfidic environments, but little is known about their metabolic potential and membership. The ability to perform dissimilatory sulfur redox would make them unique among Bacteroidetes, which are primarily known as heterotrophs that specialize in degrading complex organic molecules. Using 16S rRNA phylogeny and analysis of single amplified genomes (SAGs) from Saanich Inlet, a seasonally hypoxic basin, we elucidate the global distribution and potential metabolic capabilities of the SB clade.

Phylogenetic analysis revealed that this clade was monophyletic and had a global distribution. It is hypothesized this clade combines heterotrophic amino acid and sugar uptake with denitrification and respiratory sulfur oxidation/polysulfide reduction. Putative genes for sulfur oxidation via polysulfide reductase (*psr*) were found in the combined genome, and phylogenetic analysis confirmed these genes were likely to be *psrABC*. A denitrification pathway was present and complete save for the absence of a gene catalyzing reduction of NO to N_2O. The SB clade possesses the potential to grow by degrading a variety of polysaccharides and peptides, and possesses both aerobic respiratory and anaerobic fermentation pathways, including the TCA cycle, acetogenesis and ethanol fermentation. The presence of aerobic and anaerobic metabolic pathways makes the clade suited to environments experiencing periodic oxygen

depletion. The inferred metabolic capabilities of the SB clade taken with their wide range suggest they are potentially important players in global nitrogen and sulfur cycling.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Dr. Jason B. Sylvan and Dr. Lisa Campbell of the Department of Oceanography and Dr. Brendan Roark of the Department of Geography.

The single cell genomic data analyzed for this thesis was provided by Professor Steven Hallam of the University of British Columbia.

All other work conducted for the thesis (or) dissertation was completed by the student independently.

Funding Sources

This work was made possible in part by a fellowship from Texas A&M University and NSF under the Louis Stokes Alliance for Minority Participation Bridge to the Doctorate Fellowship. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the TAMU or the NSF.

TABLE OF CONTENTS

INTRODUCTION

One of the most ubiquitous phyla of bacteria in the ocean is the Bacteroidetes: gram-negative, rod-shaped, heterotrophic bacteria that are known to degrade complex organic matter. There is evidence that they play a major role in recycling the complex organic molecules that comprise the majority of dissolved organic matter in the ocean (Cottrell and Kirchman 2000). For example, the class Flavobacteria are a major contributor to the degradation of bacterial exopolysaccharides (Zhang et al. 2015) and members of the genera *Flavobacterium* and *Cytophaga* are able to utilize chitin, proteins, and algal polysaccharides (Cottrell and Kirchman 2000; Mann et al. 2013; Ficko-Blean et al. 2017). *Flavobacterium* and *Cytophaga* have been shown to make up a large proportion of bacteria found on marine snow (DeLong et al. 1993; Rath et al. 1998; Ploug et al. 1999) and are closely associated with microalgal blooms (Riemann et al. 2000; Fandino et al. 2001; Gomez-Pereira et al. 2010), highlighting their importance in the marine carbon cycle.

Analysis of Bacteroidetes genomes has revealed various genes that make their lifestyle possible. Bacteroidetes possess many glycoside hydrolases, glycosyl transferases, and peptidases (Bauer et al. 2006, Xie et al. 2007; Gonzalez et al. 2008; Fernández-Gómez et al 2013, Mann et al. 2013, Tang et al. 2017), which are enzymes that confer the ability to degrade complex sugars, build external adhesion structures, and degrade proteins, respectively. The hallmark of Bacteroidetes' metabolism is their polysaccharide utilization loci (PUL) - clusters of genes coding for proteins dedicated to searching out, cleaving, and taking in sugars such as cellulose, chitin, mannan, starch, xylan, pectin, and carrageenans (Xie et al. 2007; Ficko-Blean et al. 2017; Tang et al. 2017). Additionally, TonB-dependent receptors (TBDRs) and saccharide-binding

proteins such as SusD are essential in allowing Bacteroidetes to sense and take up glycans (Tang et al. 2012; Kabisch et al. 2014). Enzymes for breaking down sugars and proteins are putatively secreted using the PorSS system (McBride et al. 2004; McBride and Zhu 2013).

A gliding motility system unique to members of the Bacteroidetes allows them to move about in search of food (McBride et al. 2004; McBride and Zhu 2013), adhesion proteins allow them to latch onto surfaces (Bauer et al. 2006; Fernández-Gómez et al. 2013), and a type VI secretion system allows them to antagonize other cells in environments composed of many different microbial species (Russell et al. 2014). All of these adaptations allow members of the Bacteroidetes phylum to compete with other bacteria and exist in large numbers in different environments where organic matter is available.

In addition, there is evidence that certain members of Bacteroidetes perform denitrification in anaerobic environments, and therefore potentially play a major role in removing biologically available nitrogen in the environment (Handley et al. 2013; Jones et al. 2013). Culture and environmental studies have shown members of Bacteroidetes possess genes for different steps in the denitrification process. Enrichment cultures of activated sludge from a municipal wastewater plant in Bourgoyen-Ossemeersen, Ghent, Belgium discovered an example of Bacteroidetes capable of denitrification (Heylen et al. 2006) and a consortium of denitrifying bacteria that included Bacteroidetes were successfully cultured in a continuous flow anaerobic bioreactor with quinoline as the sole carbon source (Wang et al. 2017). Phylogenetic analyses of Bacteroidetes show they have genes for reducing nitric and nitrous oxide in anaerobic and microaerobic environments (Jones et al. 2013). The *nosZ* gene, important to N_2O reduction, has been detected in Bacteroidetes and has been found in abundance in different environmental samples from soil, wetlands, and activated sludge from Sweden, Finland, and Vietnam (Jones et

al. 2013). However, while a role for Bacteroidetes in the breakdown of complex DOM and denitrification is well established, their role in sulfidic environments and potential role in sulfur redox chemistry is underexplored.

Sulfidic environments exist in the oceans at active and inactive hydrothermal vents along mid-ocean ridges, in seafloor sediments, whale falls, and oxygen minimum zones. Numerous lineages of bacteria are able to oxidize sulfur compounds in the environment to gain energy, and in turn their metabolic activities have a significant impact on ocean biogeochemistry. For example, Espilonproteobacteria are abundant sulfur oxidizers in hydrothermal environments on vent chimney structures and underlying sediments (Sievert et al. 2008; Teske et al. 2000), particularly the genera *Sulfurospirillum*, *Sulfurovum* and *Sulfurimonas* (Campbell et al. 2006). Recently, the role of the SUP05 group of Gammaproteobacteria (Sunamura et al. 2004) has become evident due to their abundance in oxygen minimum zones (Glaubitz et al. 2013; Walsh et al. 2009), hydrothermal plumes (German et al. 2010; Lesniewski et al. 2012; Sunamura et al. 2004) and on inactive hydrothermal sulfides (Sylvan et al. 2013). Their genomes were shown to contain the sulfide dehydrogenase (*fcc*), dissimilatory sulfite reductase (*dsr*), adenylylsulfate reductase (*apr*), *Sox* and rhodanese sulfurtransferase pathways for oxidizing sulfur, and recently the first representative from the SUP05 clade was isolated and shown to grow autotrophically by oxidizing sulfur and reducing nitrate (Shah et al., 2017). This bacterium was cultured from an OMZ in the northeast Pacific Ocean. Additionally, it is possible that marine Betaproteobacteria are capable of sulfur oxidation, like their terrestrial relatives (Arsene-Ploetze et al. 2010; Chen et al. 2009; Moreira and Amils 1997) and members of Roseobacter in Alphaproteobacteria are capable of dissimilatory thiosulfate oxidation as well (Gonzalez et al. 2003).

Several sulfur oxidation pathways exist, including the *Sox* pathway, sulfite-to-sulfate oxidation, *dsr*, and polysulfide reductase (*psr*). The sox operon is a suite of proteins that is ubiquitous (Ghosh and Dam 2009) and highly variable between bacterial species. It confers the ability to oxidize various sulfur-containing substances such as thiosulfate, sulfite, and elemental sulfur (Friedrich et al. 2001). Sulfite-to-sulfate oxidation is accomplished through either a direct pathway catalyzed by a sulfite:acceptor oxidoreductase or an indirect pathway using the enzymes adenylylsulfate (APS) reductase and ATP sulfurylase and/or adenylylsulfate:phosphate adenylyltransferase with APS as an intermediate (Kappler and Dahl 2001). The *dsr* operon is the probable mechanism of dissimilatory sulfur respiration in Espsilonprotebacteria (Sievert et al. 2008) and green sulfur bacteria and, depending on the species, catalyzed by the enzymes adenosine-5'-phosphosulfate reductase or polysulfide reductase-like complex 3 (Gregersen et al. 2011). The dsr operon operates in reverse in sulfur oxidizers; it is responsible for sulfate reduction lineages capable of that metabolism. Lastly, the polysulfide reductase complex, which consists of the genes *psrABC*, catalyzes the quinone-coupled reduction of polysulfides, compounds composed of chains of sulfur atoms (Krafft et al. 1995; Jormakka et al. 2008).

Sulfur-oxidizing genes have been detected in members of Bacteroidetes, making it possible they play a previously overlooked role in marine sulfur oxidation. According to a search of the Integrated Microbial Genomes database (Markowitz et al. 2006), two marine species in the Bacteroidetes family Flavobacteriaceae, *Arenibacter certesii* and *A. latericius*, contain *soxB* in their genome. The genomes of the aquatic species *Gillisia* sp. JM1, isolated from a subglacial lake, and *Gillisia* sp. CAL575, isolated from an Antarctic sponge, contain the *soxB* gene as well. Further, two Gillisia isolates from deep-sea sediments were shown to oxidize thiosulfate, and a closely related strain was cultivated on autotrophic thiosulfate media using a sample of

subglacial brine flows from Antarctica (Mikucki and Priscu 2007). *Petrimonas sulfuriphila* while not a sulfur-oxidizer, is a terrestrial Bacteroidetes that has been shown to grow by reducing elemental sulfur to sulfide and reducing nitrate to ammonium (Grabowski et al. 2005). All of these studies point towards the possibility that Bacteroidetes play a role in sulfur redox chemistry, but none of them followed up on initial observations, leaving the question open.

Recently, a new clade of Bacteroidetes was proposed based on environmental 16S rRNA sequences from inactive hydrothermal sulfides, oxygen minimum zones, and other sulfidic environments (Sylvan et al. 2013). Sequences from this clade were first detected on hydrothermal vent chimney caps (Reysenbach et al. 2000), and later found to be abundant on inactive hydrothermal sulfides (Sylvan et al. 2012; Sylvan et al. 2013). The 16S rRNA sequences recovered from vent environments have <90% similarity to cultured representatives in the NCBI database. Further investigation found closely related 16S rRNA sequences from environmental studies of other environments, including active hydrothermal vents, sediments and marine suboxic zones, all sulfidic environments, but an absence of related sequences in non-sulfidic environments (Sylvan et al. 2013). These related 16S rRNA sequences formed a monophyletic clade and were named the Sulphiphilic Bacteroidetes (SB) clade by Sylvan and co-authors. Their presence nearly exclusively in sulfidic environments raises the question of whether the organisms represented by these 16S rRNA sequences were actively participating in sulfur redox reactions.

To answer this question, it is necessary to utilize environmental genomics, as this clade remains uncultured. In recent years, single cell genomics has enabled the in-depth analysis of individual environmental microbial species that cannot be cultured (e.g. Rinke et al. 2013), including several Bacteroidetes genomes (Woyke et al. 2009; Woyke et al. 2010). One strength

of this method is the ability to screen many samples and cells for lineages of interest that are then targeted for genomic sequencing and analysis. Here we present a global survey of 16S rRNA genes from the SB clade and an analysis of 18 single amplified genomes (SAGs) taken from Saanich Inlet, a seasonally anoxic basin in British Columbia (Zaikova et al. 2010; Torres-Beltrán et al. 2017), where bacteria capable of dissimilatory sulfur redox reactions (Wright et al. 2014; Louca et al. 2016; Hawley et al. 2017) have been found. To illustrate their potential role in their native habitat, we detail the metabolic features of the SB clade relevant to organic matter degradation, sulfur redox cycling, and denitrification.

METHODS

SAG assembly and analysis

Water samples were collected from Saanich Inlet on 09 August 2011 with a niskin bottle rosette according to the methods described in Zaikova et al. (2010), from depths of 100, 150, and 180 m. Fluorescence-activated cell sorting and sorting, multiple displacement amplification (MDA), and taxonomic analysis were performed at the Bigelow Laboratory Single Cell Genomics Center (SCGC) according to procedures described in Stepanauskas and Sieracki (2007) and Swan et al. (2013). Taxonomic identity of the SAGS was determined by sequencing the small subunit ribosomal RNA (SSU rRNA) according the methods described in Roux et al. (2014). Out of 315 SAGs (Roux et al. 2014), 18 were selected for genomic analysis because they fell with the SB clade, while a single SAG fell right outside the SB clade and was selected as an outgroup SAG. Shotgun sequencing of amplification products was done at the DOE Joint Genome Institute (JGI). MDA product was subjected to a combination of 454 and Illumina pyrosequencing to produce fragments of ~100 bp.

Alignment and binning

The raw reads in fastq format were quality-checked and assembled into draft genomes. For each SAG, interleaved raw read fastq files were split using a Perl script. Fastqc (Andrews 2010) was used to assess the quality of the reads to choose the best read orientation to use for further analysis. Trimmomatic (Bolger et al. 2014) was used to trim reads that fell below 20% quality using paired-end mode and a sliding window of 4:20. After these trimmed reads were checked with Fastqc, the forward paired reads were selected for assembly. All 19 trimmed SAGs

were assembled together using Spades (Bankevich et al. 2012) using the paired-end forward settings.

Vizbin (Lazny et al. 2015) was used to separate the resulting contigs >1000 bp into operational taxonomic units (OTUs). The Saanich Inlet contigs were combined with genomes from JGI's Integrated Microbial Genomes (IMG) database (Markowitz et al. 2012) to create a color-coded map of bins. Contigs that clustered together were considered to be part of the same bin. Contigs from the SAGs comprising each bin were extracted for further analysis.

Raw reads from the 19 SAGs were then aligned to each bin using BWA (Lee and Durbin 2009; http://bio-bwa.sourceforge.net) in order to ascertain coverage per SAG within each of the bins. The resulting alignment files were further processed using Samtools (Li et al. 2009) to make them suitable for coverage analysis. CheckM (Parks et al. 2015), a program designed to assess genome completeness, was used to calculate the percent overlap of reads between the bins and the alignments. The *merge* command was used to assess whether some of the bins should be merged, along with what SAGs could be combined to increase the amount of coverage. Bins that had a contamination of <10% when combined were merged.

Genome annotation

The assembled SAGs were inputted into the RAST database (Aziz 2008; Overbeek et al. 2014; Brettin et al. 2015) for annotation to identify functional genes of interest. A metabolic model was constructed by cross checking RAST annotations against the KEGG enzyme database (Kanehisa et al. 2016). As Bacteroidetes have been noted for their number of peptidases, we looked at potential extracellular peptidases contained within the genomic bins. To determine which peptidases were encoded by the draft genomes, RAST annotations for the 3 bins were searched for hits containing "peptidase", "protease" and "proteinase". The amino acid sequences

of the probable peptidases were checked against the MEROPS (Rawlings et al. 2016) database using HMMER (Finn et al. 2015) in order to determine their identities. HMMER and SignalP (Peterson et al. 2011) were used to determine whether the amino acid sequences contained signal peptides, which would indicate export outside of the cell membrane (Rapoport 2007). A list was created of peptidases likely to be involved in extracellular protein degradation. Peptidases which contained signal peptides but appeared to code for peptidases not involved in breakdown of extracellular proteins were eliminated from the list. Bins were run through dbCAN (Yin et al. 2012) to detect carbohydrate metabolism-related genes. Results with e values $< 1^{-10}$ and coverage >50% were selected as positive matches.

To identify Type IX Secretion System (T9SS) and gliding motility proteins within the bins, a BLAST (Altschul et al. 2005) search of Saanich Inlet Bin 1 and Bin 2 was performed against a list of T9SS proteins from *Porphyromonas gingivalis* (Lasica et al. 2017) and a list of *gld*, *spr*, and *por* proteins from *Flavobacterium johnsoniae* (McBride and Zhu 2013). Amino acid sequences for *P. gingivalis* and *F. johnsoniae* were retrieved from were retrieved from the Uniprot database. If there was a result with an e value $<e^{-10}$ or if the RAST annotation matched a member gene, it was considered that the protein in question was present.

Phylogenetic tree creation

For the 16S tree of all clade members, 16S rRNA sequences were taken off NCBI Genbank (Benson et al. 2005) based on 94% similarity to known SB clade members. The sequences were aligned using MEGA (Kumar et al. 2016) using the Clustal algorithm, and edited with Seqotron (Fourment 2016). Then the sequences were organized into OTUs through Mothur (Schloss et al. 2009) by 98% similarity, in order to remove duplicates.

After a substitution model was chosen from a best-fit test of different models, MEGA was used to create a maximum likelihood (ML) and neighbor-joining (NJ) trees using the General Time Reversible formula and rates among sites set to gamma-distributed with invariant sites, with bootstrap values assigned from 500 iterations. The NJ tree was used to verify the branches of the ML tree. The resulting ML tree was edited using Figtree (Rambaut 2014).

Using the same alignment, model fit, and tree construction methods as the 16S rRNA tree, a tree for *psrA* was constructed using homologous molybdopterin oxidoreductase catalytic subunits taken from Tully et. al (2018). Candidates for the other trees were selected by BLASTing the respective proteins of *Wolinella succinogenes*, which is known to have a working *psrABC* complex, against the Metacyc (Caspi et al. 2016), Uniprot , and NCBI databases in order to find homologous proteins and consulting existing literature for homologous proteins from organisms with known sulfur or nitrogen-reducing capabilities, including sequences used in Wright et al. (2014). Additional sequences for *psrABC* from an uncultured Saanich Inlet Marinimicrobia draft genome (Hawley et al. 2017) were obtained through the JGI IMG database. Trees were constructed with MEGA using the WAG model, gamma with invariant sites, with 500 bootstraps and a very weak branch swap filter.

Due to the limitations of constructing phylogenies using only 16S rRNA sequences, it was necessary to construct a concatenated ribosomal RNA tree (Wu and Eisen 2008) to determine the SB clade's place within Bacteroidetes. Species were selected based on a tree of Bacteroidetes members from Hahnke et al. (2016) with known *psrABC*-containing bacteria as outmembers. The 16 ribosomal proteins used are the same as in Hug et al. (2016): L2, L3, L4, L6, L14, L15, L16, L18, L22, L24, S3, S8, S10, S17, and S19. Protein sequences were taken from the Uniprot database, and each protein was aligned and trimmed separately in MEGA using

ClustalW. The resulting alignments were concatenated and a maximum likelihood tree constructed using the WAG model, gamma with invariant sites and 500 bootstraps.

RESULTS

Global distribution of the SB clade

A search of the NCBI database for 16S rRNA sequences belonging to SB clade members reveals the diversity within the SB clade (Fig. 1). As noted in earlier work (Sylvan et al. 2013), the clade is monophyletic. OTUs, defined at the 98% similarity level, showed a trend of grouping by habitat. For example, all SB clade SAGS used in this study are found on one branch of the SB clade along with other OTUs recovered from Saanich Inlet and other OMZs. The most closely related 16S rRNA sequences to the Saanich Inlet sequences and SAGs were recovered from the water column in the Atlantic Ocean and Nitinat Lake, which is an anaerobic basin (Schmidtova et al. 2009). OTUs from inactive hydrothermal sulfides, active hydrothermal vents and seeps fall onto 4 major branches of the tree, while hydrocarbon-associated habitats, including OTUs recovered from the Prestige Oil Spill and methane seeps at Hydrate Ridge and Eel River Basin grouped into two branches. One SAG, AB-746 B08AB-901, is closely related to the SB clade but serves as an outmember, falling just outside it.

A compiled map of locations where SB clade members have been detected reveals they are globally distributed (Fig. 2). Potentially due to limitations associated with collecting physical samples, most locations are near land or at known hydrothermal sites. Because availability of data is subject to ability to go to sampling sites, this map represents the minimum distribution of the clade and it is likely they are even more widespread in nature.

Concatenated rRNA tree

Concatenated rRNA phylogeny shows the Saanich Inlet SB clade is more closely related to other members of Bacteroidetes than outgroup bacteria that share *psrABC* (Fig. 3). The SB

clade was placed on a branch with members of *Draconibacterium, Odoribacter, Alistipes, Parabacteroidetes, Dysgonomonas, Porphyromonas, Barnesiella, Paraprevotella, Alloprevotella, and Prevotella*. Within this branch, the clade formed its own group with strong (93%) bootstrap support.

SAG assembly and binning

SAGs ranged from 0.17 to 1.55 Mbp in length and 3.88% to 60.00% completeness, with very low (<2%) levels of contamination/redundancy (Table 1). Three bins resulted from binning of the contigs assembled from co-assembly of all 19 SAGs. The three bins ranged from in 0.027 to 3.33 Mbp in length and 0 to 8.28% contamination/redundancy. The Saanich Inlet SAGs were grouped into three bins (Fig. 4). Aside from bin 1, which had contamination estimates of <8.28%, the bins had contamination estimates of 0% (Table 1). Out of the 3 bins, bin 3 was the smallest, 0.027 bp in size. RAST Annotation of Bin 3 revealed no genes related to central carbon metabolism, sulfur oxidation, protein and polymer degradation, or denitrification.

Coverage for each of the SAGs to the three bins is summarized in Table 2. For bins 1 and 2, the SAG was identified as either bin depending on which bin had more mapped reads. Due to its smaller size, the presence of mapped reads for bin 3 was considered a positive identification. Out of 19 SAGs, 18 mapped to bin 1, the ingroup, and one mapped to bin 2, the outgroup. Reads mapping to bin 3 were found in 4 SAGs, and all of them mapped to bin 1 as well. Therefore, it appears that bin 3 is composed of sequences which are part of an SB clade member genome.

Central carbon metabolism

Annotation of individual SAGs and bins 1-3 was performed with a focus on central carbon metabolism, degradation of DOM, denitrification and sulfur cycling (Table 3). In bin 1, hereafter referred to as the SB clade metagenome assembled genome (MAG), complete

pathways were found for glycolysis, acetogenesis, phosphoenolpyruvate conversions, the pentose phosphate cycle, glyoxylate cycle, methyglyoxal conversion, alanine conversion, and the TCA cycle. Genes for a complete glycolysis and TCA cycle imply that SB clade members can metabolize glucose in aerobic conditions and use it in the TCA cycle. A complete acetogenesis from pyruvate pathway was present as well, with 2 different potential pathways leading from pyruvate to acetyl-coA represented: pyruvate dehydrogenase and pyruvate flavodoxin oxidoreductase. Complete methylglyoxal to pyruvate and methylglyoxal to D-lactate conversion pathways were also present. Potentially, SB clade members convert sugar to pyruvate through either the methylglyoxal or glycolysis pathways and then convert the pyruvate to acetyl-coA to utilize in the TCA cycle.

The bacteria represented by the SB clade MAG appears to additionally be capable of converting various sugars. Bin 1 contained genes for glycerol kinase (*glpK*) and glycerol-3-phosphate dehydrogenase (*GPD*), suggesting the SB clade has the potential to convert glycerol to glycerone phosphate, which can be utilized in glycolysis. Two enzymes involved in the catabolism of D-ribose were present in bin 1, deoxyribose-phosphate aldolase (*deoC*) and ribokinase (*deoK*). This is similar to the D-ribose catabolism pathway of *Synechococcus elongatus*, in which 2-deoxy-D-ribose is converted into D-glyceraldehyde 3-phosphate for use in glycolysis (Caspi et al. 2016). Genes for utilizing lactase and converting mannose into β-D-Fructose-6p shed further light on metabolic options for these bacteria.

Hydrolases and transferases

The SB clade MAG has many protein sequences dedicated to glycosyl hydrolases and glycosyl transferases (Table 4), enzymes that serve as accessories to central carbon metabolism. In all, 65 glycoside hydrolases in 20 families were present, pointing to the ability to cleave a

wide variety of saccharides. In addition, the SB clade MAG possesses 68 glycosyl transferases. Annotation through dbCAN could not resolve the specific substrates of each of these genes.

Peptidases

The results of peptidase annotation of the bins through SignalP and the MEROPs database are shown in Table 5. Out of 69 putative peptidases in the SB clade MAG, 32 were predicted to be exported outside the cell, and 31 of those are likely peptidases. No extracellular peptidases were found in bin 3. The specific function of these peptidases could not be resolved through MEROPs, although RAST predicted some identities for a number of them.

Fermentation pathways

Two fermentation pathways are potentially utilized by the organism represented by the SB clade MAG; butanoate metabolism and acetogenesis linked to ethanol degradation. A pathway from acetyl-coA to butanol is complete save for the absence of an enzyme that converts crotonoyl-coA to butanoyl-coA, while butanoate and butanoyl phosphate interconversion is another pathway that is present and has potential for generating ATP. The potential for acetogenesis, and alcohol dehydrogenase (ADH) can be used to convert ethanol to acetate.

Membrane transport and motility

Various membrane proteins associated with Bacteroidetes were found in the Saanich Inlet MAG. Sequences coding for the starch-binding outer membrane proteins *susC* and *susD* were present in the genome annotation. *SusC* belongs to a class of proteins called TonB-dependent receptors (TBDRs); and is used as a marker for Bacteroidetes polysaccharide utilization loci (PULs; Tang et al. 2012). Sixteen SAGs and two bins possessed genes for TBDRs, and these proteins could serve the same purpose as *susC* for various PULs. Out of 26 Type IX Secretion System (T9SS; McBride et al. 2013) and gliding motility system (Mcbride and Zhu 2013)

proteins, bin 1 possessed 26 homologues with e values < e-^{10} (Table 6). Out of these, 2 had e values of 0 and 6 were identified by name by RAST, meaning a strong positive identification. Thus, it appears that bin 1 has a complete T9SS and gliding motility system, used in excreting proteins and cell movement.

Dissimilatory sulfur metabolism

Four genes were discovered in the SB clade MAG that are believed to take part in dissimilatory sulfur oxidation – *psrABC* and thiosulfate reductase. To verify this assumption, a phylogenomic approach was used (Eisen 1998). Genes for *psrABC* components were detected in 5 out of 19 SAGs, and the complete complex was present in one SAG and the SB clade MAG (Table 3).

Maximum likelihood trees grouped the putative *psrABC* genes with *psrABC* or homologous genes from organisms known to perform sulfur redox (Figure 5). The putative *psrA* gene grouped with confirmed *psrA* genes from *Geobacter lovleyi, Salmonella typhirium, Shewanella oneidensis, Thermus thermophilus, Wolinella succinogenes,* and a Marinomicrobia SAG with a 100% bootstrap confidence value (Figure 5A). On the *psrB* tree, the putative *psrB* gene grouped most closely with that same Marinimicrobia SAG and *psrB* from *T. thermophilus* (figure 5B). The phylogenomics of *psrC* tree is less well resolved. The putative *psrC* again grouped with another *psrC* from the same Marinimicroia SAG (Figure 5C), and the same branch contained *psrC* from *T. thermophilus* and *G. lovleyi*, but with weak bootstrap values (<50%). *psrC* and *nrfD* appear to have structural similarities and do not resolve on their own branches Indeed, sequences from *Thiohalomonas denitrificans* and *Sulfuricella denitrificans* are annotated as *nrfD/psrC*. However, taken together, these trees support the scenario that three genes in the

Saanich Inlet SB clade MAG are *psrABC*, conferring the ability to perform dissimilatory sulfur redox metabolism.

A gene annotated as "thiosulfate sulfurtransferase, rhodanese" (Table 4) by RAST was found in two SAGs and the SB clade MAG. If functional, it would reduce thiosulfate to sulfite using cyanide. While this gene is ubiquitous in living organisms, there are thiosulfate sulfurtransferases with rhodanese domains that are exclusive to *Desulfuricella* and used for sulfur respiration (Florentino et al. 2017), and it is possible the SB clade or certain species within it uses a respiratory thiosulfate sulfurtransferase.

Denitrification

In addition to sulfur cycling, OMZs are known for steep nitrate gradients caused by denitrifiers (Ulloa et al. 2012). During periods of hypoxia in Saanich Inlet, denitrification is coupled with sulfur oxidation in the sulfide–nitrate transition zone at the redoxcline between the oxygenated upper depths and hypoxic bottom water (Louca et al. 2016). Therefore we searched for and found nitrogen cycling genes in the 19 SAGs and bins 1 and 2 (Table 3, Figure 4). The organism represented by the SB clade MAG appears to have almost all the genes for complete denitrification. The nitrate reductase complex NarGHI was present, as well as the additional genes *narK* and *narJ*, which are involved in nitrite transport (Clegg et al. 2002) and Nar complex formation (Dubourdieu and DeMoss 1992). Nitrite reductase (*nirK*), which catalyzes the conversion of nitrite to nitric oxide, was detected in the SB clade MAG. A sequence coding for nitric oxide reductase, which converts NO to N_2O, was not detected in any of the SAGs, but nitrous oxide reductase (*nosZ*), which converts N_2O to N_2, was present.

Analysis of bin 2

Bin 2 is represented by a SAG just outside the SB clade (Figure 1) and is used here to discern between features present in the SB clade but not in other closely related Bacteroidetes. Because percent completion for this SAG and bin is 30.18 %, the absence of features does not necessarily mean they are absent from the organism. However, it does provide a basis for comparison.

Bin 2 lacks many central metabolism pathways present in bin 1 (table 3). The TCA cycle, which was detected in the SB clade group, was not detected in bin 2, nor were glycolysis or the methylglyoxal pathway. Alanine conversion pathways were present, but glyoxylate-related pathways were absent. The pentose pathway and acetogenesis were detected in bin 2. The PEP pathway was detected in AB-746 B08AB-901, the SAG corresponding to bin 2, but was not detected in bin 2 itself. Acetyl-coA to butanoate fermentation was present, but ethanol fermentation was absent. It is likely the lack of glycolysis or methylglyoxal pathway is a result of the lower percent completeness of the genome, as these are very common pathways that produce pyruvate that is used in acetogenesis (Kanehisa et al. 2016).

Far fewer sugar utilization pathways are present in bin 2 compared to bin 1. While D-ribose utilization and catabolism were present, lactate metabolism was absent. Mannose and glycerol utilization pathways were present in AB-746 B08AB-901 but not bin 2. While not a sugar, glycerol can be converted for use in glycolysis; glycerol pathways were not present in bin 2.

In general, fewer membrane-associated and extracellular enzymes were present in bin 2. Seven glycoside hydrolases and 17 glycosyl transferases were predicted, far fewer than in bin 1 (Table 4). The ratio of genes per Mbp of genome is smaller than the SB clade MAG as well. Out

of 20 putative peptidases in bin 2, 13 appeared to have the role of degrading extracellular proteins (Table 5). Bin 2 possessed far fewer T9SS and *gld* genes compared to bin 1. *GldA, gldG, gldF, gldI, porW/sprE, porX, porY, ompA,* a *tonB* outer membrane protein, and *porZ* appear to be present (Table 6), though whether they constitute a functional T9SS and gliding motility system is inconclusive. The genes *gldK, gldL, gldM,* and *sprA,* which are essential to gliding motility in *F. johnsoniae* (Shrivastava et al. 2013), were absent, but this could easily be due to the incompleteness of the genome. The starch-binding proteins *susC* and *susD* were absent, although TBDRs were found.

The lack of *psrABC* or other genes related to dissimilatory sulfur redox is notable, meaning that what these sulfur cycling-relating genes could be what distinguishes the SB clade from closely related Bacteroidetes from the same environment. Denitrification genes were not present either, but whether lack of both of these pathways is a product of the lower completion percentage or a true absence in this organism is unknown.

DISCUSSION

The cosmopolitan phylum Bacteroidetes is ubiquitous in the ocean and marine Bacteroidetes are known primarily as degraders of high molecular weight organic matter. They are not known to be capable of sulfur oxidation, a widespread process among marine bacteria. The discovery of closely related Bacteroidetes associated with sulfidic environments and the existence of sulfur-oxidation genes across members of the phylum prompted investigation into the lifestyle of the Sulfiphilic Bacteroidetes. Through 16S rRNA phylogeny, it was determined that the SB clade are a monophyletic branch of Bacteroidetes, and are found globally in sulfidic environments such as hydrothermal vents, methane seeps, and the redoxclines of oxygen minimum zones. Through concatenated ribosomal RNA phylogeny, the SB clade was shown to be its own branch within Bacteroidetes. Analysis of single amplified genomes from Saanich Inlet, it was determined that members of this clade possess genes for the degradation of extracellular polysaccharides and proteins and pathways for utilizing multiple monosaccharides, as well as certain membrane proteins associated with Bacteroidetes motility and secretion. They possessed both aerobic and anaerobic metabolic pathways, including the TCA cycle, acetogenesis, and ethanol fermentation. Additionally, members of this clade possess denitrification genes and a polysulfide reductase complex.

The TCA cycle is found in some form in almost all aerobic organisms and results in higher ATP yields than fermentation (Caspi et al. 2016). Given that the middle and bottom waters of Saanich Inlet are intermittently oxygenated (Zaikova et al. 2010) and the TCA cycle is common in other members of Bacteroidetes (Handley et al. 2013), its presence in the SB clade was expected. It is possible that SB clade bacteria produce oxaloacetate from the PEP to

oxaloacetate pathway and use it in the TCA cycle. Since many of the enzymes involved in the TCA cycle and glycolysis are also involved in the glyoxylate cycle, such as malate dehydrogenase (Mdh), Enolase (EnO), and citrate synthase (GltA), it is possible the SB clade possesses a functional glyoxylate cycle as well, as annotations determined the SB clade MAG possessed a complete glyoxlyate cycle.

The Saanich Inlet SB clade appears capable of using many different organic molecules in its metabolism. The presence of several saccharide-utilization pathways leading into glycolysis and the pentose phosphate pathway, plus a glycerol utilization pathway also leading into glycolysis, suggest the ability utilize diverse heterotrophic energy sources. This is further corroborated by the 65 glycoside hydrolases in the Saanich Inlet MAG, comparable to other Bacteroidetes specialized in polysaccharide and oligosaccharide degradation (Bauer et al. 2006; Tang et al. 2017). The ratio of glycoside hydrolases (19.5 per megabase pairs across 3.3 Mbp) in Bin 1 is comparable to that of *G. forsetii* and other Bacteroidetes that specialize in polysaccharides (Bauer et al. 2006), suggesting the SB clade feeds primarily on polysaccharides. Bacteroidetes are observed to have more peptidases per Mbp of genome on average than other bacteria (Fernandez-Gomez et al. 2013), and the SB clade MAG had a comparable (32) amount of extracellular peptidases for a Bacteroidetes genome of its size (25-35 peptidases per Mbp). Compared to Bin 1, Bin 2 had a lower ratio of peptidases (~28.5 peptidases per Mbp) than the SB clade MAG, but still falls within the previously measured values for Bacteroidetes (Bauer et al 2006).

In Bacteroidetes, TBDRs are hypothesized to give the phylum a competitive advantage in taking in nutrients; for example, they may specialize in secreting polysaccharide hydrolases and taking up algal sugars (Tang et al. 2012). The TBDR *susC* works in tandem with *susD* to bind

polysaccharides like starch so they can be cleaved by glycoside hydrolases. In this type of system, polysaccharides are cleaved into oligosaccharides and transported into the periplasm to be further cleaved into monosaccharides. The collection of genes that are responsible for this process are called Polysaccharide Utilization Loci (PULs; Grondin et al. 2017) and are found across marine Bacteroidetes (Kabisch et al. 2014; Barbeyron et al. 2016; Tang et al. 2017). In *G. forsetii*, *susC*-like receptors and *susD*-like binding molecules have shown to be expressed in the absence of substrate, meaning they could be used to sense sugars (Kabisch et al. 2014). It is possible *susC* and *susD* in the SB clade serve the same purpose and comprise components of PULs. However, it is uncertain whether this organism utilizes these proteins for this purpose, as *susC* and *susD*-like proteins are not required to transport cellulose across the outer membrane in *Cytophaga hutchinsii* (Zhu et al. 2015). Other TBDRs were found in the SB clade MAG and would likely serve a similar purpose to *susC* in transporting saccharides.

Gliding motility allows Bacteroidetes such as *F. johnsoniae* to move rapidly across surfaces such as agar without the use of flagella or pili. Although the exact mechanism of gliding motility is still under investigation, it is believed to be a biological motor that rotates extracellular filamentous adhesion proteins that are excreted by the cell (McBride and Zhu 2013; Shrivastava et al 2013). The MAG contained all the genes essential to gliding motility in *F. johnsoniae* (Table 6) according to McBride and Zhu (2013). Genes that were not annotated as components of these systems by RAST had BLAST results with low ($<e^{-10}$) e values, which meant they were likely matches for the missing gliding motility genes. Given that gliding motility is well documented in Bacteroidetes (McBride 2004; McBride and Zhu 2013), it is likely that these genes are present and functional. Additionally, it is possible the SB clade uses *gldG* and other *gld* genes for purposes other than motility; *gldD* and *gldG* are known to be

essential to adhesion, biofilm formation, and proteolysis in *Flavobacterium psychrophilum* (Pérez-Pascual et al. 2017).

Components of the gliding motility system are part of the *porSS*/type IX secretion system (Sato et al. 2010; McBride et al. 2013), which play a role in various functions of Bacteroidetes. In *F. johnsoniae*, the *porSS* system consists of proteins *gldK*, *gldL*, *glM*, *gldN*, *sprA*, *sprE*, and *sprT* (Sato et al. 2010; Rhodes et al. 2011). Type IX secretion systems are essential to Bacteroidetes gliding motility (Chang et al. 1984; Shrivastava et al. 2013) and also responsible for secreting chitinase in *F. johnsoniae* (Kharade and McBride 2014) and bestow virulence through secretion of protein-degrading enzymes in *Porphyromonas gingivalis* (Sato et al. 2010, Sato et al. 2013). It is possible that in addition to assisting in gliding motility, the T9SS in the SB clade is used to secrete chitinase, as several genes in the MAG were annotated as potential chitinases (Fig. 5; Table 4). Since chitin is one of the most abundant biomolecules on Earth (Muzzarelli 1999), this represents an important food source for these bacteria.

In conjunction with all the genes required to have a functional gliding motility system, the SB clade MAG contained a large number (68) of glycosyl transferases, comparable to polysaccharide specialist *G. forsetii*. Glycosyl transferases are hypothesized to be used by *G. forsetii* to synthesize cell wall components (Bauer et al. 2006) such as extracellular polysaccharide structures, which are known to be involved in surface adhesion (Hall-Stoodley et al. 2004), an adaptation suited for a lifestyle of attaching and feeding on high molecular weight organic matter.

Unlike some species of marine Bacteroidetes that have been described, the Saanich Inlet SB clade does not possess proteorhodopsin. Given that the role of bacterial proteorhodopsin is generally believed to be providing supplemental energy in the nutrient-poor surface ocean

(Gónzalez et al. 2008; Woyke et al. 2009, Fernández-Gómez et al 2013), it makes sense that this clade would not use such a process. The SAGs came from depths of >100 m, where irradiance is low (~2 W/m2 of photosynthetically active radiation; Torres-Beltran et al. 2017). In general, other locations where SB clade members have been found are also low or no light (Fig. 2).

Wolinella succinogenes is a member of the Epsilonproteobacteria known to utilize polysulfide reductase, an enzyme with 3 subunits, *psrABC*, that can reduce polysulfide or oxidize sulfide (Kraft et al. 1995). It is an integral membrane protein complex that is responsible for quinone-coupled reduction of polysulfide (Kraft et al. 1995; Jormakka et al. 2008). The *psrABC* complex can catalyze either polysulfide reduction or sulfur oxidation, as both processes have been demonstrated by *W. succinogenes psrABC (*Kraft et al. 1995; Hedderich et al. 1998). We believe that the SB clade possesses a *psr complex* because the genes were found in members of the Marinimicrobia (formerly known as Marine Group A Bacteria) from Saanich Inlet that have been posited as utilizing *psrABC* for sulfur respiration based on genetic and transcriptome evidence (Wright et al. 2014; Hawley et al. 2017). We hypothesize that the *psrABC* in the Saanich Inlet SB clade is involved in sulfur oxidation, based on geochemical modeling of Saanich Inlet which gives evidence that the rate of sulfur oxidation is far larger than the rate of sulfur reduction within the sulfide–nitrate transition zone (Louca et al. 2016). Phylogenetic trees give evidence that the putative *psrABC* in Saanich Inlet is in fact polysulfide reductase (Fig. 6). Few if any Bacteroidetes are known to perform dissimilatory sulfur oxidation. With the possibility that the SB clade has a functioning *psrABC* complex, it represents a novel group within the Bacteroidetes in terms of lifestyle.

The presence of denitrification genes in SB clade agrees with previous observations that many genera of Bacteroidetes are capable of various steps in the denitrification process.

Phylogenetic studies show there are numerous species possessing *nosZ*, part of the nitrous oxide reductase complex (Liu et al. 2008; Jung et al. 2013), as well as nitrate reductase (Mann et al. 2013) and/or nitrite reductase (NirK) (Heylen et al 2006). A study of a freshwater aquifer in Rifle, Colorado found Bacteroidetes that were believed to couple the reduction of nitrogen species with acetate oxidation via the TCA cycle as well as possessing formate-dependent nitrite/polysulfide reductase (Handley et al. 2013).

For facultative anaerobes, denitrification is an adaptation for environments where oxygen levels fluctuate and aerobic denitrification can occur when oxygen levels are low, but not zero (Ji et al. 2015). As the part of Saanich Inlet where these bacteria were taken from experiences seasonal hypoxia (Zaikova et al. 2010; Torres-Beltrán et al. 2017), this aligns with previous observations that aerobic denitrifiers can exist where oxygen levels are low.

The simultaneous ability to denitrify and oxidize sulfur would place the SB clade among many other genera of bacteria capable of sulfur redox and denitrification, including *Thiohalomonas denitrificans*, a Gammaprotebacteria found in hypersaline lakes which is capable of complete denitrification and grows with thiosulfate (Sorokin et al. 2007), and hydrothermal-vent Epsilonproteobacteria which are able to oxidize sulfur and denitrify (Sorokin et al 2004; Perez-Rodriguez et al 2013; Vetriani et al. 2014). The single cultured example of the SUP05 Gammaproteobacteria clade that is abundant in OMZs (CITE - maybe Walsh et al., 2009) and hydrothermal plumes (Sunamura et al 2004) possesses both sulfur oxidation and denitrification genes, as well (Shah et al. 2017).

It can be surmised from this analysis that the lifestyle of the Saanich Inlet SB clade is characterized by a combination of heterotrophic amino acid and sugar uptake combined with denitrification and respiratory sulfur oxidation/polysulfide reduction (Fig. 5). Using external

glycoside hydrolases and polysaccharide lyases, these bacteria cleave polysaccharides and export the resulting oligosaccharides into the periplasm using *susC* or other TBDRs, where they are further degraded by glycoside hydrolases without loss to the environment or competing bacteria. The monosaccharide components are exported into the cytoplasm where they are converted to a usable form, if necessary, and utilized in either glycolysis or the pentose phosphate pathway. The pyruvate produced through glycolysis, PEP, or alanine conversion is converted to acetyl-coA and used for various metabolic processes. In a similar manner, these bacteria export peptidases outside the cell membrane to degrade peptides and utilize the component amino acids for cellular processes. Due to the presence of essential gliding motility proteins and a T9SS, it is likely the SB clade is capable of gliding motility, moving in search of organic matter to uptake and adhering to surfaces.

The central metabolic pathway that the SB clade utilizes depends on the availability of oxygen. When conditions allow, it uses the more efficient TCA cycle to generate ATP. But when oxygen is depleted in the water, it utilizes acetogenesis, the ethanol fermentation pathway, butanoate fermentation, sulfur oxidation and denitrification to gain energy. It is this versatility that likely explains their wide presence in sulfidic environments with varying oxygen concentrations.

It is still uncertain whether the *psrAB*C of the SB clade is functional. To begin to answer this question, it will be useful to look for SB clade *psrABC* in environmental transcriptomes. This is a technique that has been successful in measuring the metabolic activity of Marinomicrobia *psrABC* in Saanich Inlet (Hawley et al. 2017). However, confirmation of sulfur redox activity will come from culture experiments where a representative of this clade is grown on various sulfur media, such as with other sulfur bacteria (Hedderich et al. 1999; Grabowski et

al. 2005; Shah et al. 2017). The genomic evidence presented here can serve as a guide for creating media targeting growth of the SB clade.

Ribosomal RNA phylogeny shows that the SB clade is globally-distributed, monophyletic, and a distinct group from other Bacteroidetes. On both the 16S and concatenated trees, the SB clade was placed on its own branch with strong (>75%) bootstrap support. Location data of 16S rRNA sequences from the SB clade shows that it is present in a wide range of sulfidic environments across the world. The approximate grouping of the 16S rRNA sequences on the maximum likelihood tree points to different members of this clade partitioning by environmental niche – active hydrothermal vent and seep sites, inactive hydrothermal sulfides, hydrocarbon-enriched sediments, and anoxic basins. The global distribution of this clade, and their quantitative contribution to sulfur and nitrogen cycling, are questions that study of environmental transcriptomes would help answer. If the SB clade is found wherever there are sulfidic environments, and truly contributing to sulfur cycling and denitrification, then it would constitute a significant player in this processes worldwide.

CONCLUSIONS

This analysis of single-single genomes from the SB clade points to a clade with the ability to perform partial denitrification and sulfur oxidation via *psrABC* and utilize a diversity of sugars in its central metabolism. The Saanich Inlet SB clade members are adapted to living in the depths of a seasonally anoxic fjord and are capable of both aerobic and anaerobic metabolisms. Although it displays many features typical of Bacteroidetes, such as a sizeable number of glycosyl transferases, glycoside hydrolases, and peptidases, the ability to perform dissimilatory sulfur oxidation distinguishes this clade as a novel lineage. The SB clade has a widespread distribution in the ocean and a potentially large role in sulfur and nitrogen cycling in the ocean, given that the sulfidic habitats in which they are abundant exist globally, and also a role in degrading high molecular weight organic matter. The global distribution of the SB clade raises the question if they are quantitatively significant to marine nitrogen and sulfur cycling. What is more, the evidence for dissimilatory sulfur redox capabilities in the SB clade and the presence of sulfur oxidation genes in other members of Bacteroidetes suggests these capabilities are potentially more common across the phylum than once thought. Investigating how this capability arose across different lineages of Bacteroidetes, and quantifying how widespread the SB is across sulfidic environments would be appropriate topics for further studies.

REFERENCES

1. **Altschul SF**, **Wootton JC**, **Gertz ME**, **Agarwala R**, **Morgulis A**, **Schäffer AA**, **Yu Y**. 2005. Protein database searches using compositionally adjusted substitution matrices. FEBS J **272**:5101–5109.
2. **Andrews S**. 2010. FastQC: a quality control tool for high throughput sequence data.
3. **Arsène-Ploetze F**, **Koechler S**, **Marchal M**, **Coppée J-Y**, **Chandler M**, **Bonnefoy V**, **Brochier-Armanet C**, **Barakat M**, **Barbe V**, **Battaglia-Brunet F**, **Bruneel O**, **Bryan CG**, **Cleiss-Arnold J**, **Cruveiller S**, **Erhardt M**, **Heinrich-Salmeron A**, **Hommais F**, **Joulian C**, **Krin E**, **Lieutaud A**, **Lièvremont D**, **Michel C**, **Muller D**, **Ortet P**, **Proux C**, **Siguier P**, **Roche D**, **Rouy Z**, **Salvignol G**, **Slyemi D**, **Talla E**, **Weiss S**, **Weissenbach J**, **Médigue C**, **Bertin PN**. 2010. Structure, Function, and Evolution of the Thiomonas spp. Genome. PLOS Genet **6**:e1000859.
4. **Aziz RK**, **Bartels D**, **Best AA**, **DeJongh M**, **Disz T**, **Edwards RA**, **Formsma K**, **Gerdes S**, **Glass EM**, **Kubal M**, **Meyer F**, **Olsen GJ**, **Olson R**, **Osterman AL**, **Overbeek RA**, **McNeil LK**, **Paarmann D**, **Paczian T**, **Parrello B**, **Pusch GD**, **Reich C**, **Stevens R**, **Vassieva O**, **Vonstein V**, **Wilke A**, **Zagnitko O**. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics **9**:75.
5. **Bankevich A**, **Nurk S**, **Antipov D**, **Gurevich AA**, **Dvorkin M**, **Kulikov AS**, **Lesin VM**, **Nikolenko SI**, **Pham S**, **Prjibelski AD**, **Pyshkin A V**, **Sirotkin A V**, **Vyahhi N**, **Tesler G**, **Alekseyev MA**, **Pevzner PA**. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol **19**:455–477.

6. **Bauer M**, **Kube M**, **Teeling H**, **Richter M**, **Lombardot T**, **Allers E**, **Würdemann CA**, **Quast C**, **Kuhl H**, **Knaust F**, **Woebken D**, **Bischof K**, **Mussmann M**, **Choudhuri J V**, **Meyer F**, **Reinhardt R**, **Amann RI**, **Glöckner FO**. 2006. Whole genome analysis of the marine Bacteroidetes"Gramella forsetii" reveals adaptations to degradation of polymeric organic matter. Environ Microbiol **8**:2201–2213.

7. **Benson DA**, **Karsch-Mizrachi I**, **Lipman DJ**, **Ostell J**, **Wheeler DL**. 2005. GenBank. Nucleic Acids Res **33**:D34–D38.

8. **Bolger AM**, **Lohse M**, **Usadel B**. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**:2114–2120.

9. **Brettin T**, **Davis JJ**, **Disz T**, **Edwards RA**, **Gerdes S**, **Olsen GJ**, **Olson R**, **Overbeek R**, **Parrello B**, **Pusch GD**, **Shukla M**, **Thomason JA**, **Stevens R**, **Vonstein V**, **Wattam AR**, **Xia F**. 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep **5**:8365.

10. **Campbell BJ**, **Engel AS**, **Porter ML**, **Takai K**. 2006. The versatile ε-proteobacteria: Key players in sulphidic habitats. Nat Rev Microbiol **4**:458–468.

11. **Caspi R**, **Altman T**, **Dreher K**, **Fulcher CA**, **Subhraveti P**, **Keseler IM**, **Kothari A**, **Krummenacker M**, **Latendresse M**, **Mueller LA**, **Ong Q**, **Paley S**, **Pujar A**, **Shearer AG**, **Travers M**, **Weerasinghe D**, **Zhang P**, **Karp PD**. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res **40**:D742–D753.

12. **Chang LE**, **Pate JL**, **Betzig RJ**. 1984. Isolation and characterization of nonspreading mutants of the gliding bacterium Cytophaga johnsonae. J Bacteriol **159**:26–35.

13. **Chen Y**, **Wu L**, **Boden R**, **Hillebrand A**, **Kumaresan D**, **Moussard H**, **Baciu M**, **Lu Y**, **Murrell JC**. 2009. Life without light: Microbial diversity and evidence of sulfur- and ammonium-based chemolithotrophy in Movile Cave. ISME J **3**:1093–1104.

14. **Clegg S**, **Yu F**, **Griffiths L**, **Cole JA**. 2002. The roles of the polytopic membrane proteins NarK, NarU and NirC in Escherichia coli K-12: two nitrate and three nitrite transporters. Mol Microbiol **44**:143–155.

15. **Cottrell MT**, **Kirchman DL**. 2000. Natural assemblages of marine proteobacteria and members of the Cytophaga-flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter. Appl Environ Microbiol **66**:1692–1697.

16. **Dubourdieu M**, **DeMoss JA**. 1992. The narJ gene product is required for biogenesis of respiratory nitrate reductase in Escherichia coli. J Bacteriol **174**:867–872.

17. **Eisen JA**. 1998. Phylogenomics: Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis. Genome Res **8** :163–167.

18. **Eisen JA**. 1998. Phylogenomics: Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis. Genome Res 163–167.

19. **Fandino LB**, **Riemann L**, **Steward GF**, **Long RA**, **Azam F**. 2001. Variations in bacterial community structure during a dinoflagellate bloom analyzed by DGGE and 16S rDNA sequencing.

20. **Fernández-Gómez B**, **Richter M**, **Schüler M**, **Pinhassi J**, **Acinas SG**, **González JM**, **Pedrós-Alió C**. 2013. Ecology of marine bacteroidetes: A comparative genomics approach. ISME J **7**:1026–1037.

21. **Ficko-Blean E**, **Préchoux A**, **Thomas F**, **Rochat T**, **Larocque R**, **Zhu Y**, **Stam M**, **Génicot S**, **Jam M**, **Calteau A**, **Viart B**, **Ropartz D**, **Pérez-Pascual D**, **Correc G**,

Matard-Mann M, **Stubbs KA**, **Rogniaux H**, **Jeudy A**, **Barbeyron T**, **Médigue C**, **Czjzek M**, **Vallenet D**, **McBride MJ**, **Duchaud E**, **Michel G**. 2017. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. Nat Commun **8**:1685.

22. **Finn RD**, **Clements J**, **Arndt W**, **Miller BL**, **Wheeler TJ**, **Schreiber F**, **Bateman A**, **Eddy SR**. 2015. HMMER web server: 2015 update. Nucleic Acids Res W30–W38.

23. **Florentino AP**, **Stams AJM**, **Sánchez-Andrea I**. 2017. Genome Sequence of Desulfurella amilsii Strain TR1 and Comparative Genomics of Desulfurellaceae Family. Front Microbiol **8**:222.

24. **Fourment M**, **Holmes EC**. 2016. Seqotron: a user-friendly sequence editor for Mac OS X. BMC Res Notes **9**:106.

25. **Friedrich CG**, **Rother D**, **Bardischewsky F**, **Quentmeier A**, **Fischer J**. 2001. Oxidation of Reduced Inorganic Sulfur Compounds by Bacteria: Emergence of a Common Mechanism? Appl Environ Microbiol **67**:2873–2882.

26. **German CR**, **Bowen A**, **Coleman ML**, **Honig DL**, **Huber JA**, **Jakuba M V.**, **Kinsey JC**, **Kurz MD**, **Leroy S**, **McDermott JM**, **de Lepinay BM**, **Nakamura K**, **Seewald JS**, **Smith JL**, **Sylva SP**, **Van Dover CL**, **Whitcomb LL**, **Yoerger DR**. 2010. Diverse styles of submarine venting on the ultraslow spreading Mid-Cayman Rise. Proc Natl Acad Sci **107**:14020–14025.

27. **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**:999–1043.

28. **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 **79** :2767–2776.

29. **Gómez-Pereira PR**, **Fuchs BM**, **Alonso C**, **Oliver MJ**, **van Beusekom JEE**, **Amann R**. 2010. Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. Isme J **4**:472.

30. **Gonzalez JM**, **Fernandez-Gomez B**, **Fernandez-Guerra A**, **Gomez-Consarnau L**, **Sanchez O**, **Coll-Llado M**, **del Campo J**, **Escudero L**, **Rodriguez-Martinez R**, **Alonso-Saez L**, **Latasa M**, **Paulsen I**, **Nedashkovskaya O**, **Lekunberri I**, **Pinhassi J**, **Pedros-Alio C**. 2008. Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). Proc Natl Acad Sci **105**:8724–8729.

31. **González JM**, **Covert JS**, **Whitman WB**, **Henriksen JR**, **Mayer F**, **Scharf B**, **Schmitt R**, **Buchan A**, **Fuhrman JA**, **Kiene RP**, **Moran MA**. 2003. Silicibacter pomeroyi sp. nov. and Roseovarius nubinhibens sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. Int J Syst Evol Microbiol **53**:1261–1269.

32. **Grabowski A**, **Tindall BJ**, **Bardin V**, **Blanchet D**, **Jeanthon C**. 2005. Petrimonas sulfuriphila gen. nov., sp. nov., a mesophilic fermentative bacterium isolated from a biodegraded oil reservoir. Int J Syst Evol Microbiol **55**:1113–1121.

33. **Gregersen LH**, **Bryant DA**, **Frigaard N-U**. 2011. Mechanisms and Evolution of Oxidative Sulfur Metabolism in Green Sulfur Bacteria. Front Microbiol **2**:1–14.

34. **Grondin JM**, **Tamura K**, **Déjean G**, **Abbott DW**, **Brumer H**. 2017. Polysaccharide Utilization Loci: Fuelling microbial communities. J Bacteriol **199**:e00860–16.

35. **Hahnke RL**, **Meier-Kolthoff JP**, **García-López M**, **Mukherjee S**, **Huntemann M**, **Ivanova NN**, **Woyke T**, **Kyrpides NC**, **Klenk HP**, **Göker M**. 2016. Genome-based taxonomic classification of Bacteroidetes. Front Microbiol **7**.

36. **Hall-Stoodley L**, **Costerton JW**, **Stoodley P**. 2004. Bacterial biofilms: from the Natural environment to infectious diseases. Nat Rev Microbiol **2**:95.

37. **Handley KM**, **Verberkmoes NC**, **Steefel CI**, **Williams KH**, **Sharon I**, **Miller CS**, **Frischkorn KR**, **Chourey K**, **Thomas BC**, **Shah MB**, **Long PE**, **Hettich RL**, **Banfield JF**. 2013. Biostimulation induces syntrophic interactions that impact C, S and N cycling in a sediment microbial community. ISME J **7**:800–816.

38. **Hawley AK**, **Nobu MK**, **Wright JJ**, **Durno WE**, **Morgan-Lang C**, **Sage B**, **Schwientek P**, **Swan BK**, **Rinke C**, **Torres-Beltrán M**, **Mewis K**, **Liu WT**, **Stepanauskas R**, **Woyke T**, **Hallam SJ**. 2017. Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients. Nat Commun **8**:1–9.

39. **Hedderich R**, **Klimmek O**, **Kroeger A**, **Dirmeier R**, **Keller M**, **Stetter KO**. 1999. Anaerobic respiration with elemental sulfur and with disulfides. FEMS Microbiol Rev **22**:353–381.

40. **Heylen K**, **Gevers D**, **Vanparys B**, **Wittebolle L**, **Geets J**, **Boon N**, **De Vos P**. 2006. The incidence of nirS and nirK and their genetic heterogeneity in cultivated denitrifiers. Environ Microbiol **8**:2012–2021.

41. **Holmes DE**, **Nevin KP**, **Woodard TL**, **Peacock AD**, **Lovley DR**. 2007. Prolixibacter bellariivorans gen. nov., sp. nov., a sugar-fermenting, psychrotolerant anaerobe of the phylum Bacteroidetes, isolated from a marine-sediment fuel cell. Int J Syst Evol Microbiol **57**:701–707.

42. **Hug LA**, **Baker BJ**, **Anantharaman K**, **Brown CT**, **Probst AJ**, **Castelle CJ**, **Butterfield CN**, **Hernsdorf AW**, **Amano Y**, **Ise K**, **Suzuki Y**, **Dudek N**, **Relman DA**, **Finstad KM**, **Amundson R**, **Thomas BC**, **Banfield JF**. 2016. A new view of the tree of life. Nat Microbiol **1**:16048.

43. **Ji B**, **Yang K**, **Zhu L**, **Jiang Y**, **Wang H**, **Zhou J**, **Zhang H**. 2015. Aerobic denitrification: A review of important advances of the last 30 years. Biotechnol Bioprocess Eng **20**:643–651.

44. **Jones CM**, **Graf DRH**, **Bru D**, **Philippot L**, **Hallin S**. 2013. The unaccounted yet abundant nitrous oxide-reducing microbial community: A potential nitrous oxide sink. ISME J **7**:417–426.

45. **Jormakka M**, **Yokoyama K**, **Yano T**, **Tamakoshi M**, **Akimoto S**, **Shimamura T**, **Curmi P**, **Iwata S**. 2008. Molecular mechanism of energy conservation in polysulfide respiration. Nat Struct Mol Biol **15**:730–737.

46. **Jung J**, **Choi S**, **Jung H**, **Scow KM**, **Park W**. 2013. Primers for amplification of nitrous oxide reductase genes associated with Firmicutes and Bacteroidetes in organic-compound-rich soils. Microbiology **159**:307–315.

47. **Kabisch A**, **Otto A**, **König S**, **Becher D**, **Albrecht D**, **Schüler M**, **Teeling H**, **Amann RI**, **Schweder T**. 2014. Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes “Gramella forsetii” KT0803. ISME J **8**:1492–1502.

48. **Kanehisa M**, **Sato Y**, **Kawashima M**, **Furumichi M**, **Tanabe M**. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res **44**:D457–D462.

49. **Kappler U**, **Dahl C**. 2001. Enzymology and molecular biology of prokaryotic sulfate oxidationFEMS Microbiology Letters.

50. **Kharade SS**, **McBride MJ**. 2014. Flavobacterium johnsoniae Chitinase ChiA Is Required for Chitin Utilization and Is Secreted by the Type IX Secretion System. J Bacteriol **196**:961–970.

51. **Kirchman DL**. 2002. The ecology of Cytophaga-Flavobacteria in aquatic environments. FEMS Microbiol Ecol **39**:91–100.

52. **Kumar S**, **Stecher G**, **Tamura K**. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol **33**:1870–1874.

53. **Laczny CC**, **Sternal T**, **Plugaru V**, **Gawron P**, **Atashpendar A**, **Margossian HH**, **Coronado S**, **der Maaten L van**, **Vlassis N**, **Wilmes P**. 2015. VizBin - an application for reference-independent visualization and human-augmented binning of metagenomic data. Microbiome **3**:1.

54. **Landry Z**, **Swa BK**, **Herndl GJ**, **Stepanauskas R**, **Giovannoni SJ**. 2017. SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. MBio **8**:1–19.

55. **Lasica AM**, **Ksiazek M**, **Madej M**, **Potempa J**. 2017. The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function . Front Cell Infect Microbiol .

56. **Lesniewski RA**, **Jain S**, **Anantharaman K**, **Schloss PD**, **Dick GJ**. 2012. The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. Isme J **6**:2257.

57. **Li H**, **Durbin R**. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**:1754–1760.

58. **Li H**, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079.

59. **Liu X**, **Gao C**, **Zhang A**, **Jin P**, **Wang L**, **Feng L**. 2008. The nos gene cluster from gram-positive bacterium Geobacillus thermodenitrificans NG80-2 and functional characterization of the recombinant NosZ. FEMS Microbiol Lett **289**:46–52.

60. **Long ED**, **Franks DG**, **Alldredge AL**. 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol Oceanogr **38**:924–934.

61. **Louca S**, **Hawley AK**, **Katsev S**, **Torres-Beltran M**, **Bhatia MP**, **Kheirandish S**, **Michiels CC**, **Capelle D**, **Lavik G**, **Doebeli M**, **Crowe SA**, **Hallam SJ**. 2016. Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. Proc Natl Acad Sci **113**:E5925 LP – E5933.

62. **Mann AJ**, **Hahnke RL**, **Huang S**, **Werner J**, **Xing P**, **Barbeyron T**, **Huettel B**, **Stüber K**, **Reinhardt R**, **Harder J**, **Glöckner FO**, **Amann RI**, **Teeling H**. 2013. The Genome of the Alga-Associated Marine Flavobacterium Formosa agariphila KMM 3901(T) Reveals a Broad Potential for Degradation of Algal Polysaccharides. Appl Environ Microbiol **79**:6813–6822.

63. **Markowitz VM**, **Chen I-MA**, **Palaniappan K**, **Chu K**, **Szeto E**, **Grechkin Y**, **Ratner A**, **Jacob B**, **Huang J**, **Williams P**, **Huntemann M**, **Anderson I**, **Mavromatis K**, **Ivanova NN**, **Kyrpides NC**. 2012. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res **40**:D115–D122.

64. **Markowitz VM**, **Korzeniewski F**, **Palaniappan K**, **Szeto E**, **Werner G**, **Padki A**, **Zhao X**, **Dubchak I**, **Hugenholtz P**, **Anderson I**, **Lykidis A**, **Mavromatis K**, **Ivanova N**,

Kyrpides NC. 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res **34**:D344–D348.

65. **McBride MJ**. 2004. Cytophaga-Flavobacterium Gliding Motility. J Mol Microbiol Biotechnol **7**:63–71.

66. **McBride MJ**, **Zhu Y**. 2013. Gliding motility and por secretion system genes are widespread among members of the phylum bacteroidetes. J Bacteriol **195**:270–278.

67. **Mikucki JA**, **Priscu JC**. 2007. Bacterial diversity associated with blood falls, a subglacial outflow from the Taylor Glacier, Antarctica. Appl Environ Microbiol **73**:4029–4039.

68. **Moreira D**, **Amils R**. 1997. Phylogeny of Thiobacillus cuprinus and other mixotrophic thiobacilli: proposal for Thiomonas gen. nov. Int J Syst Bacteriol **47**:522–528.

69. **Muzzarelli RA**. 1999. Native, industrial and fossil chitins. EXS **87**:1–6.

70. **Newton RJ**, **Griffin LE**, **Bowles KM**, **Meile C**, **Gifford S**, **Givens CE**, **Howard EC**, **King E**, **Oakley CA**, **Reisch CR**, **Rinta-Kanto JM**, **Sharma S**, **Sun S**, **Varaljay V**, **Vila-Costa M**, **Westrich JR**, **Moran MA**. 2010. Genome characteristics of a generalist marine bacterial lineage. ISME J **4**:784–798.

71. **Overbeek R**, **Olson R**, **Pusch GD**, **Olsen GJ**, **Davis JJ**, **Disz T**, **Edwards RA**, **Gerdes S**, **Parrello B**, **Shukla M**, **Vonstein V**, **Wattam AR**, **Xia F**, **Stevens R**. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res **42**:D206–D214.

72. **Parks DH**, **Imelfort M**, **Skennerton CT**, **Hugenholtz P**, **Tyson GW**. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res **25** :1043–1055.

73. **Pérez-Pascual D**, **Rochat T**, **Kerouault B**, **Gómez E**, **Neulat-Ripoll F**, **Henry C**, **Quillet E**, **Guijarro JA**, **Bernardet JF**, **Duchaud E**. 2017. More Than Gliding: Involvement of GldD and GldG in the Virulence of Flavobacterium psychrophilum. Front Microbiol **8**:2168.

74. **Petersen TN**, **Brunak S**, **von Heijne G**, **Nielsen H**. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods **8**:785.

75. **Ploug H**, **Grossart H-P**, **Azam F**, **Jørgensen B**, **Jørgensen**. 1999. Photosynthesis, respiration, and carbon turnover in sinking marine snow from surface waters of Southern California Bight: Implications for the carbon cycle in the ocean. Mar Ecol Prog Ser **79**:1–11.

76. **Rambaut A**. 2014. FigTree. Mol Evol Phylogenetics Epidemiol.

77. **Rapoport TA**. 2007. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature **450**:663.

78. **Rath J**, **Wu KY**, **Herndl G**, **DeLong EF**. 1998. High phylogenetic diversity in a marine-snow-associated bacterial assemblage. Aquat Microb Ecol - AQUAT MICROB ECOL **14**:261–269.

79. **Rawlings ND**, **Barrett AJ**, **Finn R**. 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res **44**:D343–D350.

80. **Reysenbach A-L**, **Longnecker K**, **Kirshtein J**. 2000. Novel Bacterial and Archaeal Lineages from an In Situ Growth Chamber Deployed at a Mid-Atlantic Ridge Hydrothermal Vent. Appl Environ Microbiol **66**:3798–3806.

81. **Rhodes RG**, **Samarasam MN**, **Van Groll EJ**, **McBride MJ**. 2011. Mutations in Flavobacterium johnsoniae sprE Result in Defects in Gliding Motility and Protein Secretion. J Bacteriol **193**:5322–5327.

82. **Riemann L**, **Steward GF**, **Azam F**. 2000. Dynamics of Bacterial Community Composition and Activity during a Mesocosm Diatom Bloom. Appl Environ Microbiol **66**:578–587.

83. **Rinke C**, **Schwientek P**, **Sczyrba A**, **Ivanova NN**, **Anderson IJ**, **Cheng J-F**, **Darling A**, **Malfatti S**, **Swan BK**, **Gies EA**, **Dodsworth JA**, **Hedlund BP**, **Tsiamis G**, **Sievert SM**, **Liu W-T**, **Eisen JA**, **Hallam SJ**, **Kyrpides NC**, **Stepanauskas R**, **Rubin EM**, **Hugenholtz P**, **Woyke T**. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature **499**:431.

84. **Russell AB**, **Wexler AG**, **Harding BN**, **Whitney JC**, **Bohn AJ**, **Goo YA**, **Tran BQ**, **Barry NA**, **Zheng H**, **Peterson SB**, **Chou S**, **Gonen T**, **Goodlett DR**, **Goodman A.**, **Mougous JD**. 2014. A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. Cell Host Microbe **16**:227–226.

85. **Ryan WBF**, **Carbotte SM**, **Coplan JO**, **O'Hara S**, **Melkonian A**, **Arko R**, **Weissel. Rose Anne**, **Ferrini V**, **Goodwillie A**, **Nitsche F**, **Bonczkowski J**, **Zemsky R**. 2009. Global Multi-Resolution Topography synthesis. Geochemistry, Geophys Geosystems **10**.

86. **Sato K**, **Naito M**, **Yukitake H**, **Hirakawa H**, **Shoji M**, **McBride MJ**, **Rhodes RG**, **Nakayama K**. 2010. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. Proc Natl Acad Sci U S A **107**:276–281.

87. **Sato K**, **Yukitake H**, **Narita Y**, **Shoji M**, **Naito M**, **Nakayama K**. 2013. Identification of Porphyromonas gingivalis proteins secreted by the Por secretion system. FEMS Microbiol Lett **338**:68–76.

88. **Schloss PD**, **Westcott SL**, **Ryabin T**, **Hall JR**, **Hartmann M**, **Hollister EB**, **Lesniewski RA**, **Oakley BB**, **Parks DH**, **Robinson CJ**, **Sahl JW**, **Stres B**, **Thallinger GG**, **Van Horn DJ**, **Weber CF**. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol **75**:7537–7541.

89. **Schmidtova J**, **Hallam SJ**, **Baldwin SA**. 2009. Phylogenetic diversity of transition and anoxic zone bacterial communities within a near-shore anoxic basin: Nitinat Lake. Environ Microbiol **11**:3233–3251.

90. **Shah V**, **Chang BX**, **Morris RM**. 2017. Cultivation of a chemoautotroph from the SUP05 clade of marine bacteria that produces nitrite and consumes ammonium. ISME J **11**:263–271.

91. **Shrivastava A**, **Johnston JJ**, **van Baaren JM**, **McBride MJ**. 2013. Flavobacterium johnsoniae GldK, GldL, GldM, and SprA Are Required for Secretion of the Cell Surface Gliding Motility Adhesins SprB and RemA. J Bacteriol **195**:3201–3212.

92. **Sievert SM**, **Scott KM**, **Klotz MG**, **Chain PSG**, **Hauser LJ**, **Hemp J**, **Hügler M**, **Land M**, **Lapidus A**, **Larimer FW**, **Lucas S**, **Malfatti SA**, **Meyer F**, **Paulsen IT**, **Ren Q**, **Simon J**, **Class the USFG**. 2008. Genome of the Epsilonproteobacterial Chemolithoautotroph Sulfurimonas denitrificans . Appl Environ Microbiol **74** :1145–1156.

93. **Sorokin DY**, **Tourova TP**, **Braker G**, **Muyzer G**. 2007. Thiohalomonas denitrificans gen. nov., sp. nov. and Thiohalomonas nitratireducens sp. nov., novel obligately chemolithoautotrophic, moderately halophilic, thiodenitrifying Gammaproteobacteria from hypersaline habitats. Int J Syst Evol Microbiol **57**:1582–1589.

94. **Stepanauskas R**, **Sieracki ME**. 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc Natl Acad Sci **104**:9052–9057.

95. **Stewart FJ**. 2011. Dissimilatory sulfur cycling in oxygen minimum zones: an emerging metagenomics perspective. Biochem Soc Trans **39**:1859–1863.

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

97. **Swan BK**, **Tupper B**, **Sczyrba A**, **Lauro FM**, **Martinez-Garcia M**, **González JM**, **Luo H**, **Wright JJ**, **Landry ZC**, **Hanson NW**, **Thompson BP**, **Poulton NJ**, **Schwientek P**, **Acinas SG**, **Giovannoni SJ**, **Moran MA**, **Hallam SJ**, **Cavicchioli R**, **Woyke T**, **Stepanauskas R**. 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci U S A **110**:11463–11468.

98. **Sylvan JB**, **Toner BM**, **Edwards KJ**. 2012. Life and Death of Deep-Sea Vents: Bacterial Diversity and Ecosystem Succession on Inactive Hydrothermal Sulfides. mBio **3** .

99. **Sylvan JB**, **Sia TY**, **Haddad AG**, **Briscoe LJ**, **Toner BM**, **Girguis PR**, **Edwards KJ**. 2013. Low Temperature Geomicrobiology Follows Host Rock Composition Along a Geochemical Gradient in Lau Basin. Front Microbiol **4**.

100. **Tang K**, **Jiao N**, **Liu K**, **Zhang Y**, **Li S**. 2012. Distribution and functions of tonb-dependent transporters in marine bacteria and environments: Implications for dissolved organic matter utilization. PLoS One **7**.

101. **Tang K**, **Lin Y**, **Han Y**, **Jiao N**. 2017. Characterization of potential polysaccharide utilization systems in the marine Bacteroidetes Gramella flava JLT2011 using a multi-omics approach. Front Microbiol **8**:1–13.

102. **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.

103. **Torres-Beltrán M**, **Hawley AK**, **Capelle D**, **Zaikova E**, **Walsh DA**, **Mueller A**, **Scofield M**, **Payne C**, **Pakhomova L**, **Kheirandish S**, **Finke J**, **Bhatia M**, **Shevchuk O**, **Gies EA**, **Fairley D**, **Michiels C**, **Suttle CA**, **Whitney F**, **Crowe SA**, **Tortell PD**, **Hallam SJ**. 2017. A compendium of geochemical information from the Saanich Inlet water column. Sci Data **4**:170159.

104. **Tully BJ**, **Wheat CG**, **Glazer BT**, **Huber JA**. 2018. A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. ISME J **12**:1–16.

105. **Ulloa O**, **Canfield DE**, **DeLong EF**, **Letelier RM**, **Stewart FJ**. 2012. Microbial oceanography of anoxic oxygen minimum zones. Proc Natl Acad Sci **109**:15996–16003.

106. **Walsh DA**, **Zaikova E**, **Howes CG**, **Song YC**, **Wright JJ**, **Tringe SG**, **Tortell PD**, **Hallam SJ**. 2009. Metagenome of a Versatile Chemolithoautotroph from Oceanic Dead Zones. Science (80-) **326**:4–6.

107. **Wang Y**, **Tian H**, **Huang F**, **Long W**, **Zhang Q**, **Wang J**, **Zhu Y**, **Wu X**, **Chen G**, **Zhao L**, **Bakken LR**, **Frostegård Å**, **Zhang X**. 2017. Time-resolved analysis of a denitrifying bacterial community revealed a core microbiome responsible for the anaerobic degradation of quinoline. Sci Rep **7**:1–11.

108. **Woyke T**, **Tighe D**, **Mavromatis K**, **Clum A**, **Copeland A**, **Schackwitz W**, **Lapidus A**, **Wu D**, **McCutcheon JP**, **McDonald BR**, **Moran NA**, **Bristow J**, **Cheng J-F**. 2010. One Bacterial Cell, One Complete Genome. PLoS One **5**:e10314.

109. **Woyke T**, **Xie G**, **Copeland A**, **González JM**, **Han C**, **Kiss H**, **Saw JH**, **Senin P**, **Yang C**, **Chatterji S**, **Cheng JF**, **Eisen JA**, **Sieracki ME**, **Stepanauskas R**. 2009. Assembling the marine metagenome, one cell at a time. PLoS One **4**:e5299.

110. **Wright JJ**, **Mewis K**, **Hanson NW**, **Konwar KM**, **Maas KR**, **Hallam SJ**. 2014. Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. ISME J **8**:455–468.

111. **Wu M**, **Eisen JA**. 2008. A simple, fast, and accurate method of phylogenomic inference. Genome Biol **9**:R151.

112. **Xie G**, **Bruce DC**, **Challacombe JF**, **Chertkov O**, **Detter JC**, **Gilna P**, **Han CS**, **Lucas S**, **Misra M**, **Myers GL**, **Richardson P**, **Tapia R**, **Thayer N**, **Thompson LS**, **Brettin TS**, **Henrissat B**, **Wilson DB**, **McBride MJ**. 2007. Genome Sequence of the Cellulolytic Gliding Bacterium Cytophaga hutchinsonii . Appl Environ Microbiol **73** :3536–3546.

113. **Yin Y**, **Mao X**, **Yang J**, **Chen X**, **Mao F**, **Xu Y**. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res **40**:W445–W451.

114. **Zaikova E**, **Walsh DA**, **Stilwell CP**, **Mohn WW**, **Tortell PD**, **Hallam SJ**. 2010. Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. Environ Microbiol **12**:172–191.

115. **Zhang Z**, **Chen Y**, **Wang R**, **Cai R**, **Fu Y**, **Jiao N**. 2015. The fate of marine bacterial exopolysaccharide in natural marine microbial communities. PLoS One **10**:1–16.

116. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res **45**:D158–D169.

APPENDIX A

FIGURES

Figure 1: Maximum likelihood tree of SB clade and outgroups, showing ≥50% bootstrap values. SB clade members are color-coded by habitat.

Figure 2: Map of locations from where members of the Sulphiphilic Bacteroidetes have been recovered, with locations color-coded by environment type. Map created in Geo Map App (Ryan et al., 2009).

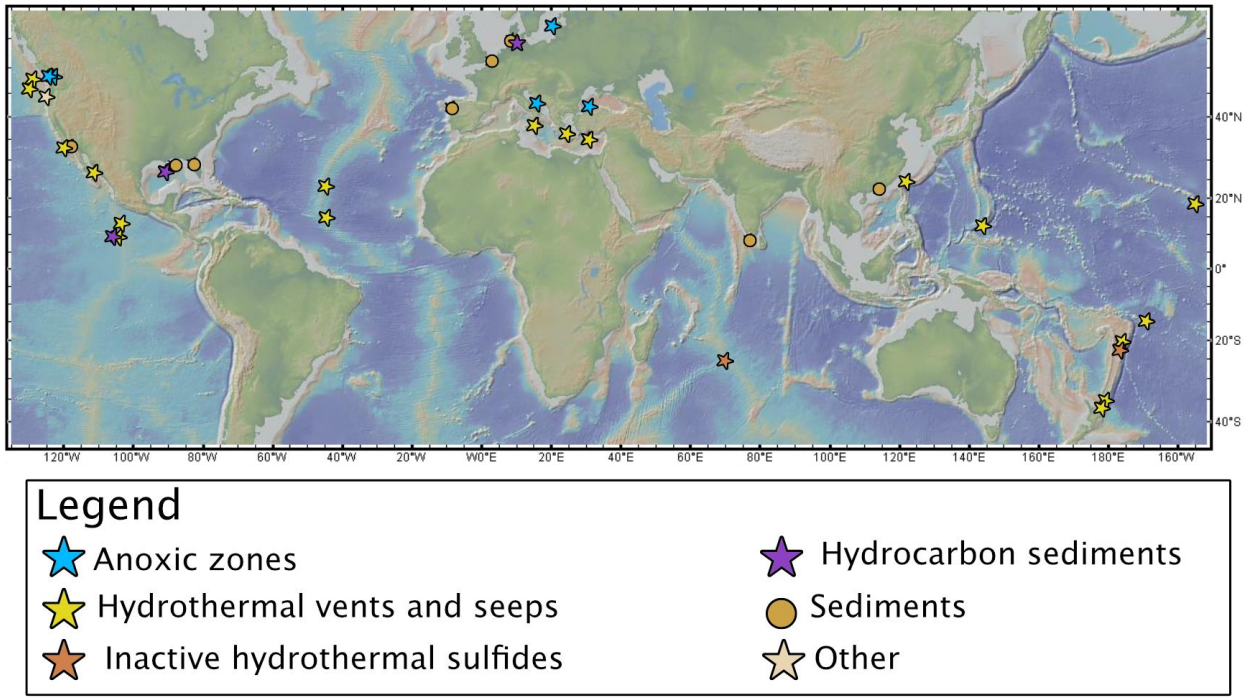


Figure 3: Concatenated rRNA tree with relative bootstrap values, showing the relationship of the SB clade to other Bacteroidetes.

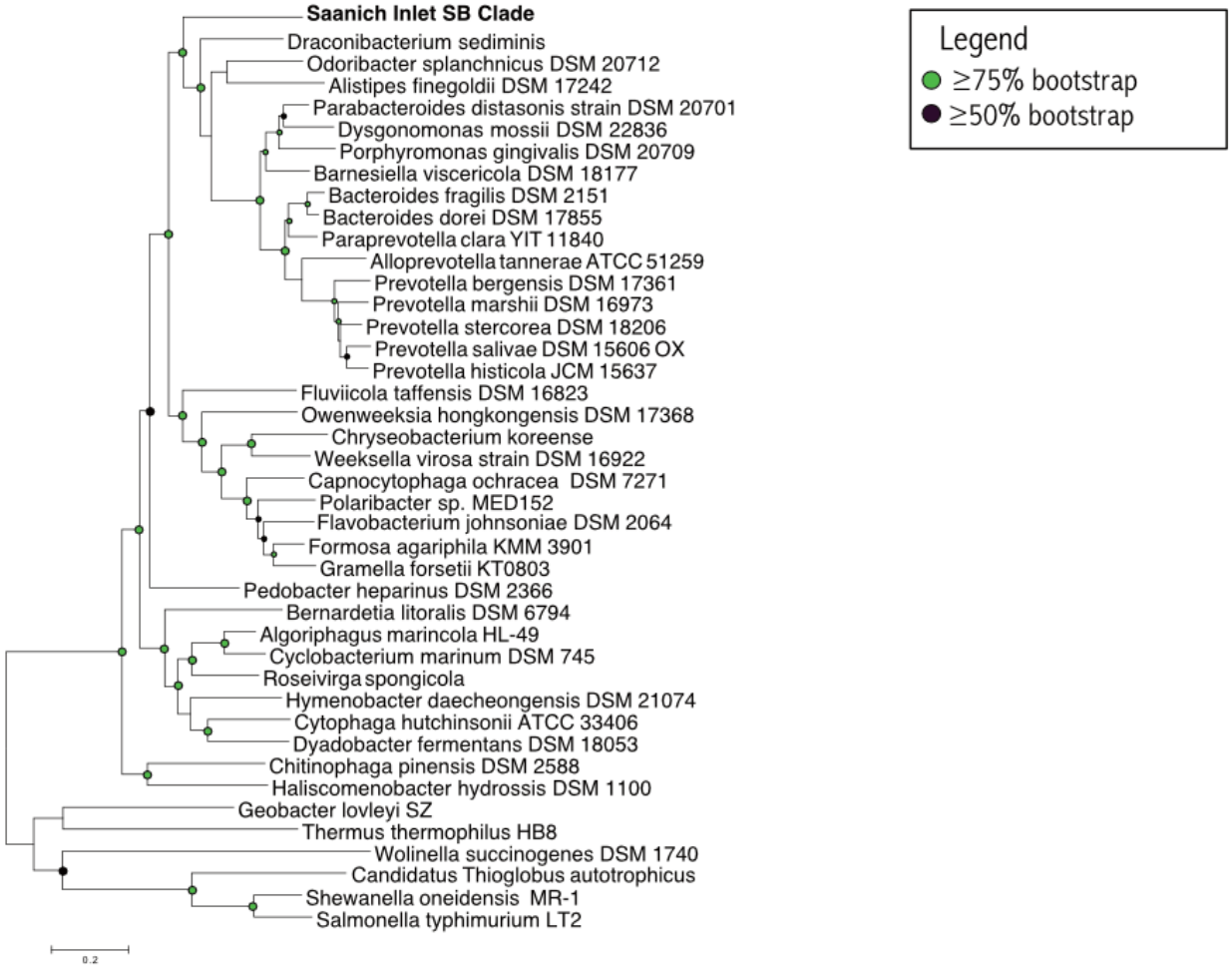


Figure 4: Vizbin map showing the grouping of the Saanich Inlet combined genome contigs and contigs of outgroup organisms from IMG. Bin 3 is not outlined, but consists of the blue dots outside bin 1 and bin 2.

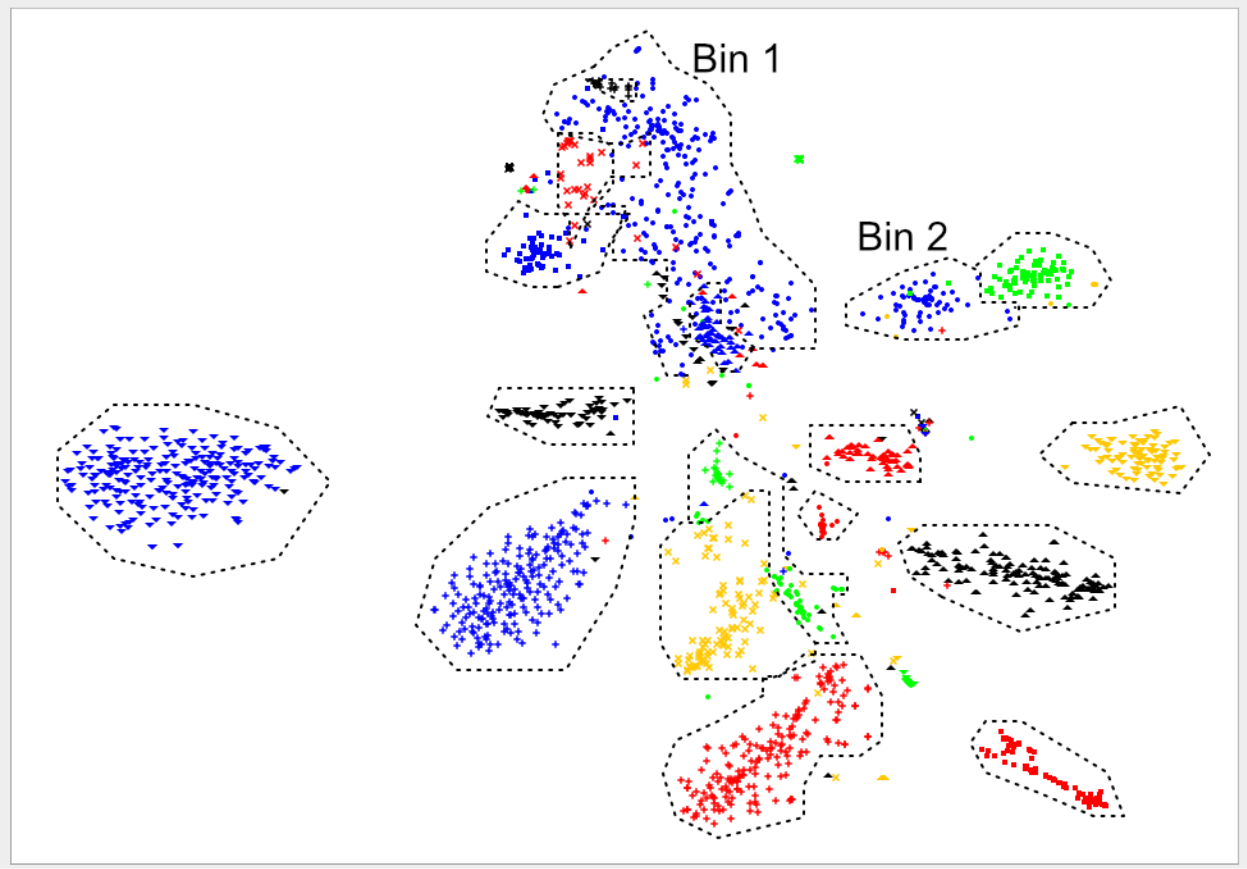


Figure 5: Metabolic model of hypothetical Saanich Inlet SB clade member showing important metabolic pathways and hypothesized location of reactions and enzymes. Dotted lines represent pathways for which the catalyzing enzyme was missing from the genome, and asterisks denote genes found in the first assembly of SAGs but not the final assembled MAG.

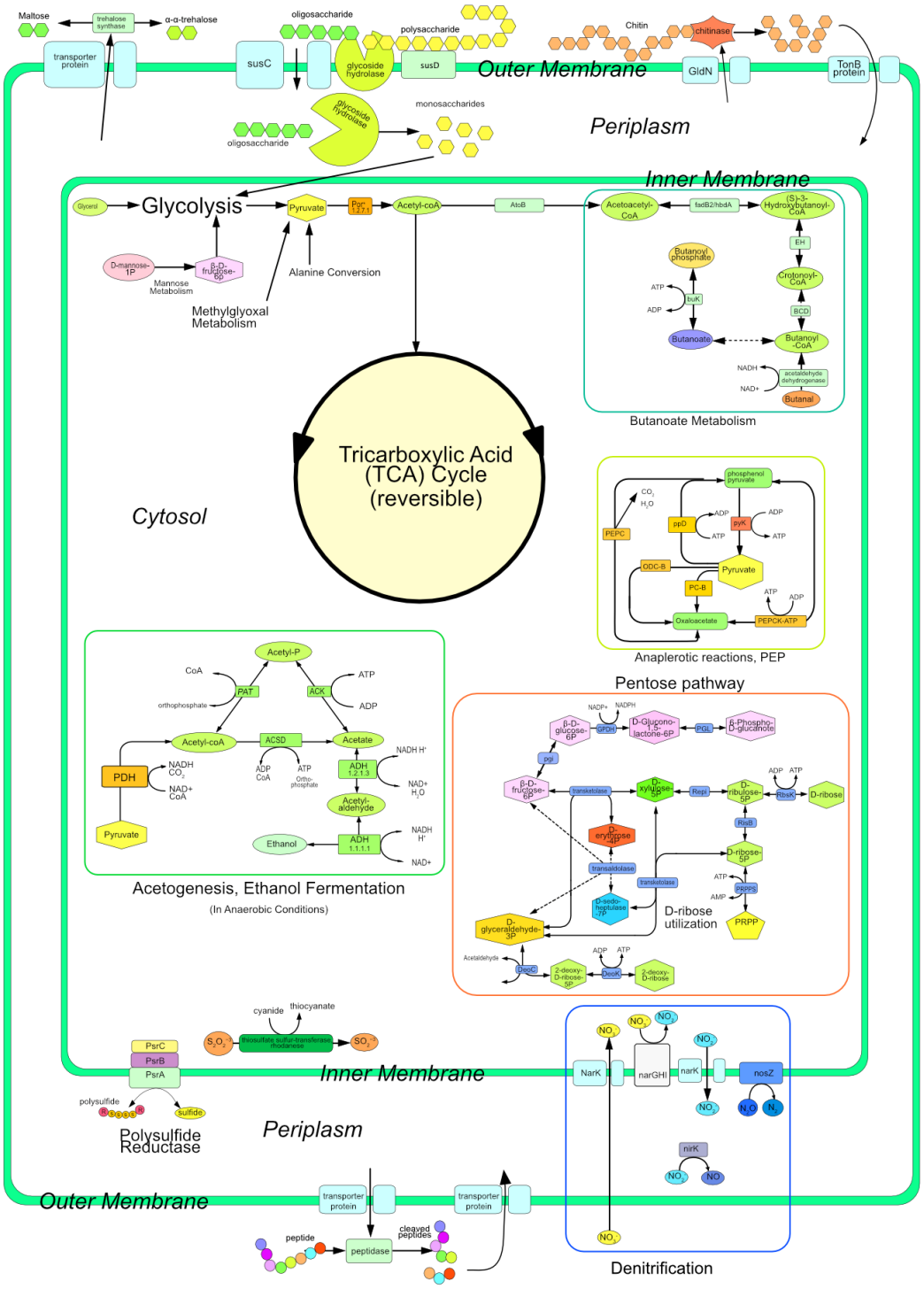
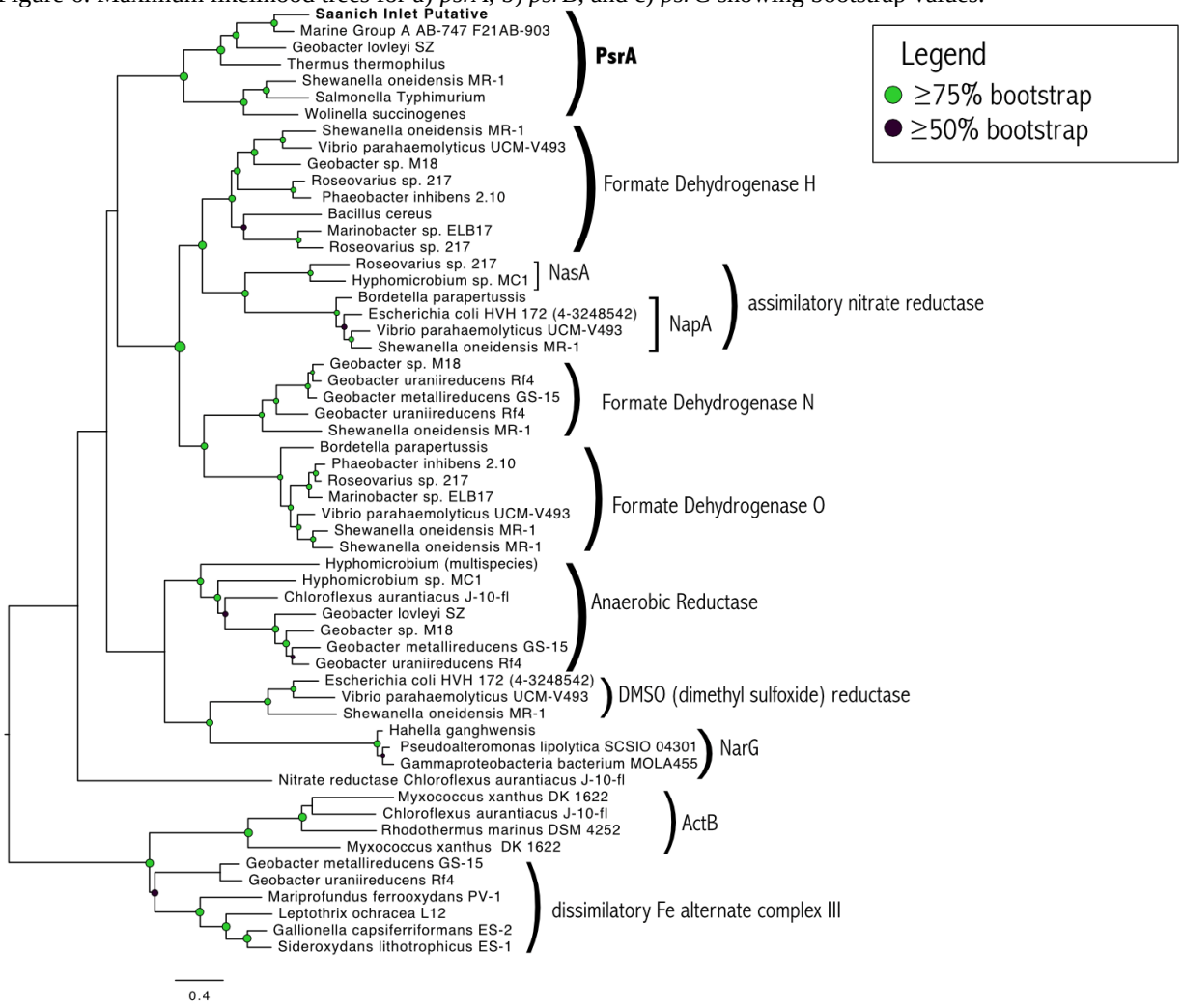
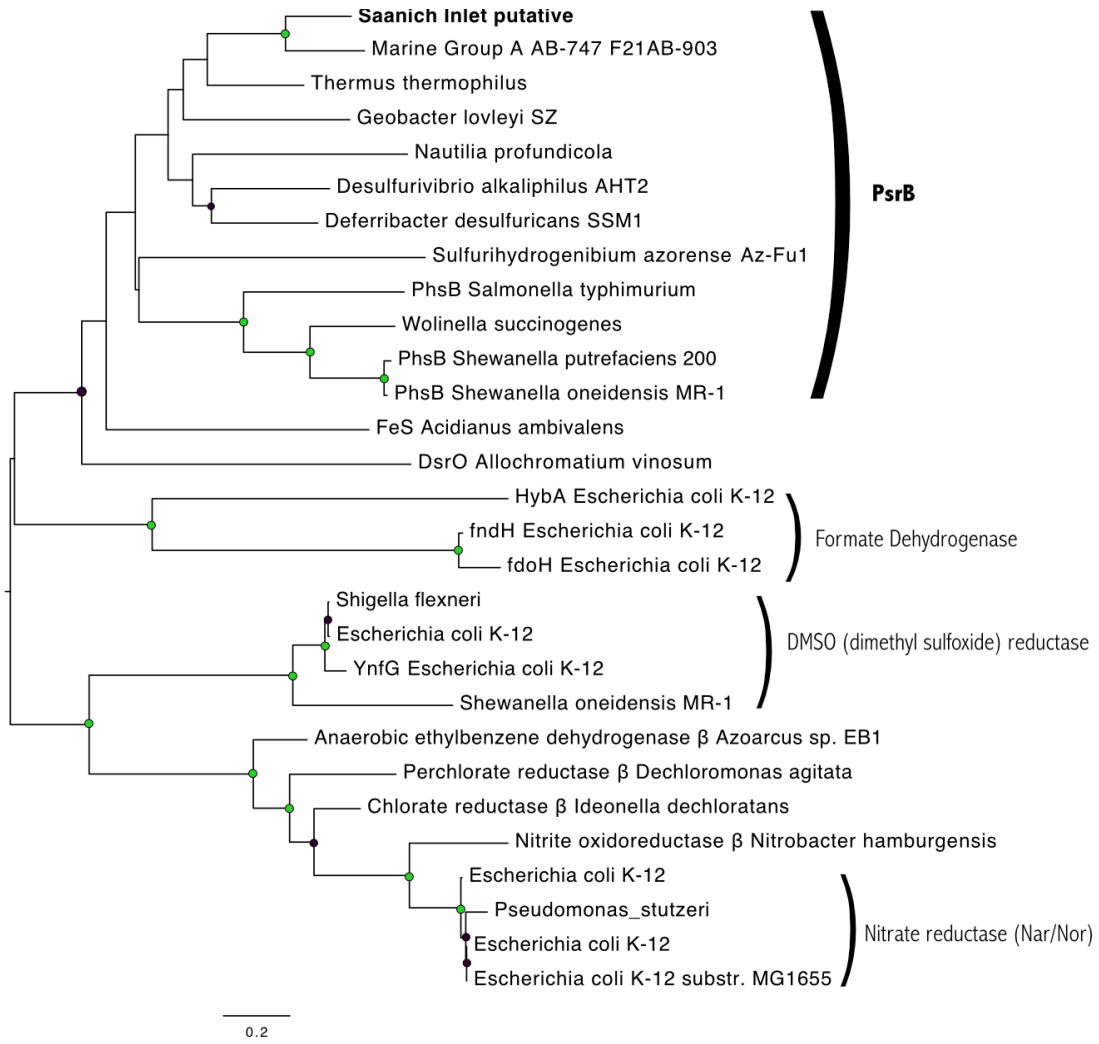
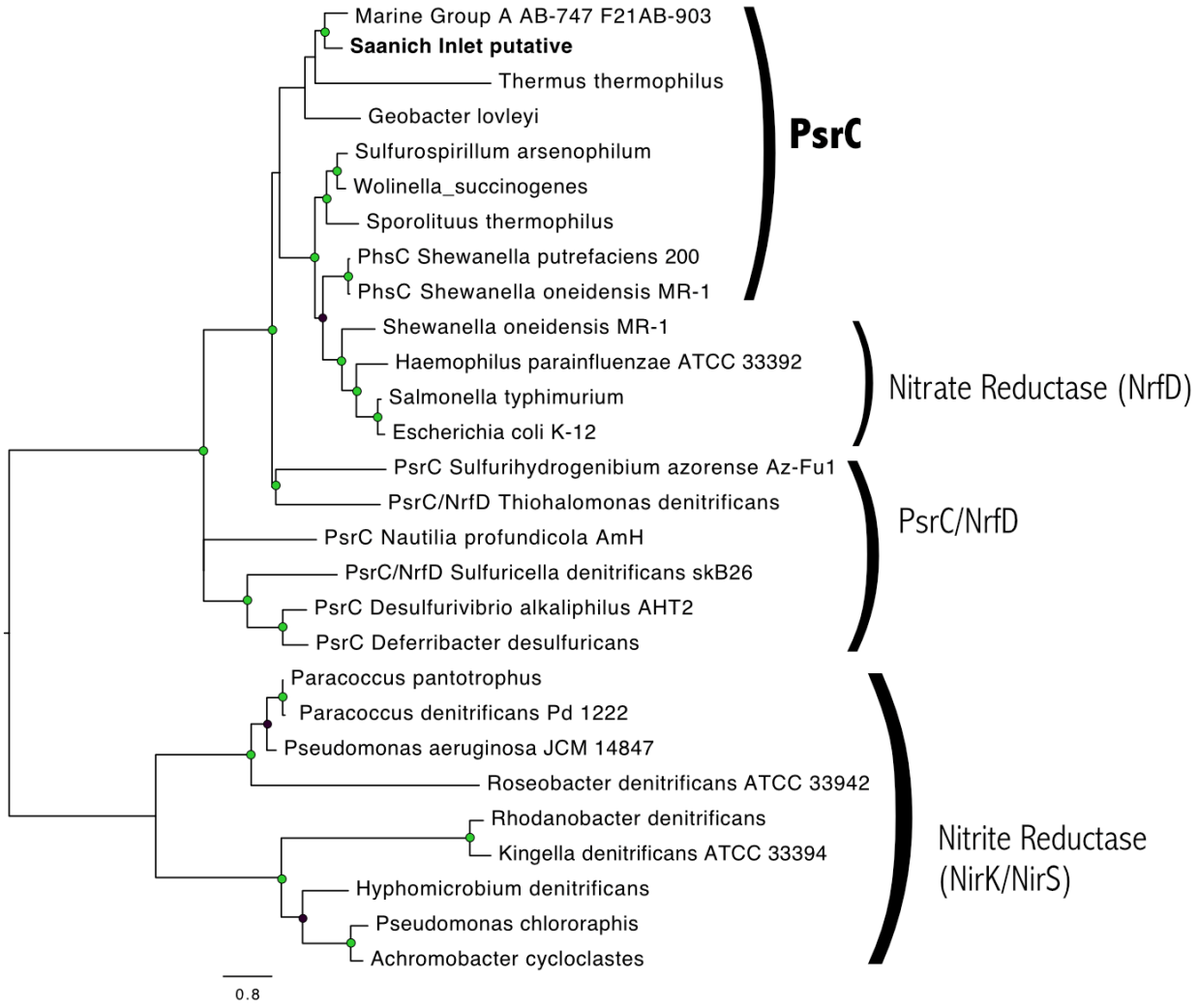


Figure 6: Maximum likelihood trees for a) *psrA*, b) *psrB*, and c) *psrC* showing bootstrap values.




Saanich Inlet putative
Marine Group A AB-747 F21AB-903
Thermus thermophilus
Geobacter lovleyi SZ
Nautilia profundicola
Desulfurivibrio alkaliphilus AHT2
Deferribacter desulfuricans SSM1
Sulfurihydrogenibium azorense Az-Fu1
PhsB Salmonella typhimurium
Wolinella succinogenes
PhsB Shewanella putrefaciens 200
PhsB Shewanella oneidensis MR-1
PsrB
FeS Acidianus ambivalens
DsrO Allochromatium vinosum
HybA Escherichia coli K-12
fndH Escherichia coli K-12
fdoH Escherichia coli K-12
Formate Dehydrogenase
Shigella flexneri
Escherichia coli K-12
YnfG Escherichia coli K-12
Shewanella oneidensis MR-1
DMSO (dimethyl sulfoxide) reductase
Anaerobic ethylbenzene dehydrogenase β Azoarcus sp. EB1
Perchlorate reductase β Dechloromonas agitata
Chlorate reductase β Ideonella dechloratans
Nitrite oxidoreductase β Nitrobacter hamburgensis
Escherichia coli K-12
Pseudomonas_stutzeri
Escherichia coli K-12
Escherichia coli K-12 substr. MG1655
Nitrate reductase (Nar/Nor)
0.2


Marine Group A AB-747 F21AB-903
Saanich Inlet putative
Thermus thermophilus
Geobacter lovleyi
Sulfurospirillum arsenophilum
Wolinella_succinogenes
Sporolituus thermophilus
PhsC Shewanella putrefaciens 200
PhsC Shewanella oneidensis MR-1
PsrC
Shewanella oneidensis MR-1
Haemophilus parainfluenzae ATCC 33392
Salmonella typhimurium
Escherichia coli K-12
Nitrate Reductase (NrfD)
PsrC Sulfurihydrogenibium azorense Az-Fu1
PsrC/NrfD Thiohalomonas denitrificans
PsrC Nautilia profundicola AmH
PsrC/NrfD Sulfuricella denitrificans skB26
PsrC Desulfurivibrio alkaliphilus AHT2
PsrC Deferribacter desulfuricans
PsrC/NrfD
Paracoccus pantotrophus
Paracoccus denitrificans Pd 1222
Pseudomonas aeruginosa JCM 14847
Roseobacter denitrificans ATCC 33942
Rhodanobacter denitrificans
Kingella denitrificans ATCC 33394
Hyphomicrobium denitrificans
Pseudomonas chlororaphis
Achromobacter cycloclastes
Nitrite Reductase
(NirK/NirS)
0.8

APPENDIX B

TABLES

Table 1: SAGs and genomic bins with statistics calculated through CheckM, with the outgroup SAG highlighted.

SAG	Assembly Length	Sample Depth (m)	N50	max contig length	# contigs	# predicted genes	% com-pleteness	% contamination
AB-746_J14AB-902	756588	100	23274	55638	65	716	24.62	0
AB-746_O23AB-902	1166311	100	19837	72845	80	1070	45.52	0.56
AB-747_I02AB-903_A04	645949	100	18051	64728	52	614	24.86	0
AB-747_J10AB-903	993960	100	35367	154838	61	915	36.11	0
AB-750B07AB-903	303989	150	15998	24250	33	316	10.22	0
AB-750C04AB-903	1201624	150	20523	64410	86	1118	42.22	0.54
AB-750C21AB-904	850983	150	32676	60934	51	763	32.51	0.15
AB-751_E20AB-904	1308066	150	37595	80952	79	1173	41.18	0
AB-751_K14AB-905	951346	150	25386	49083	70	863	36.67	1.09
AB-754_A03AB-905	1160621	185	24911	61468	76	1065	44.14	0
AB-754_A14AB-905	1550734	185	37859	113503	77	1365	60.99	0
AB-755_D07E10	905487	185	17835	55354	76	853	39.14	0.54
AB-755_F03B01	973479	185	26906	88439	72	888	31.82	0
AB-755_K22C05	528898	185	13760	36016	54	503	17.65	0.02
AB-755_M23E07	790602	185	18968	63705	70	722	36.83	0
AB-755_N03F07	1154223	185	25632	57959	75	1051	50.86	0.54
AB-755_N04G07	171357	185	10720	28129	20	205	3.88	0
AB-755_P17C10	864288	185	28992	53923	55	796	31.9	0.86
AB-746_B08AB-901	745353	100	25447	52135	63	776	28.94	0
BIN 1	3325997		51607	217125	283	3033	95.14	8.28
BIN 2	706393		33666	57487	58	727	30.18	0
BIN 3	27218		3086	3594	8	27	0	0

Table 2: Coverage status for each SAG, by bin.

SAG	Bin 1	Bin 2	Bin 3
AB-746_B08AB-901	No	Yes	No
AB-755_F03B01	Yes	No	No
AB-755_P17C10	Yes	No	No
AB-750B07AB-903	Yes	No	No
AB-751_E20AB-904	Yes	No	No
AB-746_J14AB-902	Yes	No	No
AB-746_O23AB-902	Yes	No	No
AB-754_A03AB-905	Yes	No	No
AB-755_N03F07	Yes	No	No
AB-750C04AB-903	Yes	No	No
AB-754_A14AB-905	Yes	No	No
AB-751_K14AB-905	Yes	No	No
AB-747_J10AB-903	Yes	No	Yes
AB-747_I02AB-903_A04	Yes	No	No
AB-755_N04G07	Yes	No	No
AB-755_K22C05	Yes	No	Yes
AB-755_D07E10	Yes	No	No
AB-755_M23E07	Yes	No	Yes
AB-750C21AB-904	Yes	No	Yes

Table 3: Metabolic pathways present in SAGs and bins. A (+) indicates RAST's prediction of a complete pathway. A (*) indicates a pathway that RAST did not annotate as complete but were manually assessed as probably complete. The psrABC operon encodes the polysulfide reductase complex (Braatsch et. al 2002).

	AB-754 A03AB-905	AB-750 B07AB-903	AB-746 J14AB-902	AB-750 C04AB-903	AB-755 N03F07	AB-751 K14AB-905	AB-751 E20AB-904	AB-750 C21AB-904	AB-755 P17C10	AB-755 M23E07	AB-755 N04G07	AB-754 A14AB-905	AB-755 D07E10	AB-755 K22C05	AB-747 J10AB-903	AB-755 F03B01	AB-747 I02AB-903	AB-746 B08AB-901	Bin 1	Bin 2
Central Carbohydrate Metabolism																				
Dehydrogenase complexes																		+	+	
Glycolate, glyoxylate interconversions							+		+						+		+		*	
Glycolysis and Gluconeogenesis			+	+			+	+				+			+		+		+	
Methylglyoxal Metabolism	+	+			+				+				+	+	+				+	
Pentose phosphate pathway				+				+	+							+		+	+	+
Pyruvate Alanine Serine Interconversions			+	+	+		+			+		+	+		+		+	+	+	+
anaplerotic reactions, PEP	+					+	+			+		+	+	+		+		+	+	
acetyl-coA, acetogenesis from pyruvate	+				+		+			+		+			+	+	+	+	+	+
Pyruvate:ferredoxin oxidoreductase	+		+	+		+	+		+					+		+	+		+	
TCA cycle																			+	
Fermentation																				
Acetyl-coA fermentation to Butyrate			+						+	+	+	+			+	+		+	+	+
Fermentations: Mixed acid					+		+												+	
Monosaccharides																				
D-ribose utilization				+		+	+								+	+	+	+	+	+
Deoxyribose & Deoxynucleoside Catabolism				+		+	+					+			+	+	+	+	+	+
Mannose Metabolism				+	+	+			+			+			+	+	+	+	+	
One-carbon Metabolism																				
One-carbon metabolism by tetrahydropterines	+		+		+	+	+	+		+		+			+	+	+		+	
Serine-glyoxylate cycle																			+	
Organic acids																				
Lactate utilization	+								+			+							+	

Table 3, continued.

Organic acids																				
Citrate Metabolism, Transport, & Regulation	+				+	+		+	+							+			+	
Cellulosome		+										+							+	
Protein Degradation																				
Aminopeptidases		+		+	+	+	+	+	+			+	+		+	+	+	+	+	+
Metallocarboxypeptidases	+		+		+		+					+	+			+	+		+	
Dipeptidases	+		+		+	+	+	+	+	+		+	+	+		+	+	+	+	+
Serine endopeptidase			+	+	+	+	+		+	+		+	+	+	+			+	+	+
Omegapeptidases																		+		+
Sugar Alcohols																				
Glycerol, Glycerol-3-phosphate Uptake & Utilization										+									+	
Nitrogen																				
Denitrification				+	+	+													+	
Denitrifying reductase gene clusters	+		+	+	+	+	+		+	+				+		+			+	
Nitrate and nitrite ammonification		+		+	+	+	+					+	+		+	+			+	
Sulfur																				
thiosulfate sulfurtransferase (rhodanese)							+					+							+	
psrA	+				+		+					+							+	
psrB	+				+		+												+	
psrC	+				+		+					+							+	
Motility and Outer Membrane																				
gld motility proteins	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+
TonB dependent receptors		+		+	+	+	+	+	+	+		+	+	+	+	+		+	+	+

Table 4. List of carbohydrate-related gene families in bins a) 1 and b) 2. Known functions of each family are listed. GH = glycoside hydrolase, GT = Glycosyl Transferase.

bin 1		
type	**#**	**example functions**
GH109	7	α-N-acetylgalactosaminidase (EC 3.2.1.49)
GH13	2	α-amylase (EC 3.2.1.1); pullulanase (EC 3.2.1.41); cyclomaltodextrin glucanotransferase (EC 2.4.1.19)
GH13_16	1	maltohexaose-forming α-amylase (AmyM) 3.2.1.98 (Corallococcus sp. EGB), STHERM_c15980 3.2.1.1 (Spirochaeta thermophila DSM 6192)
GH13_20	1	cyclomaltodextrinase (3.2.1.54), cyclic maltosyl-maltose hydrolase (3.2.1.-), maltogenic α-amylase (3.2.1.54)
GH142	1	β-L-arabinofuranosidase (EC 3.2.1.185)
GH16	4	xyloglucan:xyloglucosyltransferase (EC 2.4.1.207); keratan-sulfate endo-1,4-β-galactosidase (EC 3.2.1.103); endo-1,3-β-glucanase (EC 3.2.1.39)
GH17	1	glucan endo-1,3-β-glucosidase (EC 3.2.1.39); glucan 1,3-β-glucosidase (EC 3.2.1.58); licheninase (EC 3.2.1.73)
GH2	1	β-galactosidase (EC 3.2.1.23) ; β-mannosidase (EC 3.2.1.25); β-glucuronidase (EC 3.2.1.31)
GH20	1	β-hexosaminidase (EC 3.2.1.52); lacto-N-biosidase (EC 3.2.1.140); β-1,6-N-acetylglucosaminidase) (EC 3.2.1.-)
GH23	3	lysozyme type G (EC 3.2.1.17); peptidoglycan lyase (EC 4.2.2.n1) a.k.a. transglycosylase; chitinase (EC 3.2.1.14)
GH3	4	β-glucosidase (EC 3.2.1.21); xylan 1,4-β-xylosidase (EC 3.2.1.37); β-glucosylceramidase (EC 3.2.1.45)
GH31	1	α-glucosidase (EC 3.2.1.20); α-galactosidase (EC 3.2.1.22); α-mannosidase (EC 3.2.1.24)
GH33	3	sialidase or neuraminidase (EC 3.2.1.18); trans-sialidase (EC 2.4.1.-); 2-keto-3-deoxynononic acid hydrolase (EC 3.2.1.-)
GH5	2	endo-β-1,4-glucanase / cellulase (EC 3.2.1.4); endo-β-1,4-xylanase (EC 3.2.1.8); β-glucosidase (EC 3.2.1.21)
GH53	1	endo-β-1,4-galactanase (EC 3.2.1.89).
GH64	1	β-1,3-glucanase (EC 3.2.1.39)
GH65	1	α,α-trehalase (EC 3.2.1.28); maltose phosphorylase (EC 2.4.1.8); trehalose phosphorylase (EC 2.4.1.64)
GH73	2	lysozyme (EC 3.2.1.17); mannosyl-glycoprotein endo-β-N-acetylglucosaminidase (EC 3.2.1.96); peptidoglycan hydrolase w/ endo-β-N-acetylglucosaminidase specificity (EC 3.2.1.-)
GH74	22	endoglucanase (EC 3.2.1.4); oligoxyloglucan reducing end-specific cellobiohydrolase (EC 3.2.1.150); xyloglucanase (EC 3.2.1.151)
GH81	1	endo-β-1,3-glucanase (EC 3.2.1.39)
GH92	3	mannosyl-oligosaccharide α-1,2-mannosidase (EC 3.2.1.113); mannosyl-oligosaccharide α-1,3-mannosidase (EC 3.2.1.-); mannosyl-oligosaccharide α-1,6-mannosidase (EC 3.2.1.-)
GH93	2	exo-α-L-1,5-arabinanase (EC 3.2.1.-)
Total		65

Table 4, continued.

bin 1		
type	**#**	**example functions**
GT11	1	GDP-L-Fuc: galactoside α-1,2-L-fucosyltransferase (EC 2.4.1.69); GDP-L-Fuc: β-LacNac α-1,3-L-fucosyltransferase (EC 2.4.1.-)
GT13	1	α-1,3-mannosyl-glycoprotein β-1,2-N-acetylglucosaminyltransferase (EC 2.4.1.101)
GT19	1	lipid-A-disaccharide synthase (EC 2.4.1.182).
GT2	1	cellulose synthase (EC 2.4.1.12); chitin synthase (EC 2.4.1.16); dolichyl-phosphate β-D-mannosyltransferase (EC 2.4.1.83)
GT22	1	Dol-P-Man: Man6GlcNAc2-PP-Dol α-1,2-mannosyltransferase (EC 2.4.1.259); Dol-P-Man: Man8GlcNAc2-PP-Dol α-1,2-mannosyltransferase (EC 2.4.1.261); Dol-P-Man: Man2-GlcNAc-phosphatidylinositol α-1,2-mannosyltransferase (EC 2.4.1.-)
GT27	1	polypeptide α-N-acetylgalactosaminyltransferase (EC 2.4.1.41)
GT28	2	1,2-diacylglycerol 3-β-galactosyltransferase (EC 2.4.1.46); 1,2-diacylglycerol 3-β-glucosyltransferase (EC 2.4.1.157); UDP-GlcNAc: Und-PP-MurAc-pentapeptide β-N-acetylglucosaminyltransferase (EC 2.4.1.227)
GT30	1	CMP-β-KDO: α-3-deoxy-D-manno-octulosonic-acid (KDO) transferase (EC 2.4.99.-).
GT39	1	Family Dol-P-Man: protein α-mannosyltransferase (EC 2.4.1.109)
GT4	49	sucrose synthase (EC 2.4.1.13); sucrose-phosphate synthase (EC 2.4.1.14); α-glucosyltransferase (EC 2.4.1.52)
GT41	1	UDP-GlcNAc: peptide β-N-acetylglucosaminyltransferase (EC 2.4.1.255); UDP-Glc: peptide N-β-glucosyltransferase (EC 2.4.1.-)
GT5	1	UDP-Glc: glycogen glucosyltransferase (EC 2.4.1.11); ADP-Glc: starch glucosyltransferase (EC 2.4.1.21); NDP-Glc: starch glucosyltransferase (EC 2.4.1.242)
GT51	2	murein polymerase (EC 2.4.1.129)
GT83	2	undecaprenyl phosphate-α-L-Ara4N: 4-amino-4-deoxy-β-L-arabinosyltransferase (EC 2.4.2.43); dodecaprenyl phosphate-β-galacturonic acid: lipopolysaccharide core α-galacturonosyl transferase (EC 2.4.1.-)
GT87	1	polyprenol-P-Man: α-1,2-mannosyltransferase (EC 2.4.1.-)
GT9	2	lipopolysaccharide N-acetylglucosaminyltransferase (EC 2.4.1.56); heptosyltransferase (EC 2.4.-.-)
Total		68

Table 4, continued.

bin 2		
type	**#**	**example functions**
GH109	1	α-N-acetylgalactosaminidase (EC 3.2.1.49)
GH18	1	chitinase (EC 3.2.1.14); lysozyme (EC 3.2.1.17); endo-β-N-acetylglucosaminidase (EC 3.2.1.96);
GH23	2	lysozyme type G (EC 3.2.1.17); peptidoglycan lyase (EC 4.2.2.n1) a.k.a. peptidoglycan lytic transglycosylase; chitinase (EC 3.2.1.14)
GH33	1	sialidase or neuraminidase (EC 3.2.1.18); trans-sialidase (EC 2.4.1.-); 2-keto-3-deoxynononic acid hydrolase (EC 3.2.1.-)
GH73	1	lysozyme (EC 3.2.1.17); mannosyl-glycoprotein endo-β-N-acetylglucosaminidase (EC 3.2.1.96); peptidoglycan hydrolase w/ endo-β-N-acetylglucosaminidase specificity (EC 3.2.1.-)
GH92	1	mannosyl-oligosaccharide α-1,2-mannosidase (EC 3.2.1.113); mannosyl-oligosaccharide α-1,3-mannosidase (EC 3.2.1.-); mannosyl-oligosaccharide α-1,6-mannosidase (EC 3.2.1.-)
Total	7	
GT2	7	cellulose synthase (EC 2.4.1.12); chitin synthase (EC 2.4.1.16); dolichyl-phosphate β-D-mannosyltransferase (EC 2.4.1.83);
GT27	1	polypeptide α-N-acetylgalactosaminyltransferase (EC 2.4.1.41)
GT4	7	sucrose synthase (EC 2.4.1.13); sucrose-phosphate synthase (EC 2.4.1.14); α-glucosyltransferase (EC 2.4.1.52);
GT83	2	undecaprenyl phosphate-α-L-Ara4N: 4-amino-4-deoxy-β-L-arabinosyltransferase (EC 2.4.2.43); dodecaprenyl phosphate-β-galacturonic acid: lipopolysaccharide core α-galacturonosyl transferase (EC 2.4.1.-)
Total		17

Table 5. Peptidases found in Bin 1 and Bin 2, with predicted identities and catalytic type.

bin 1		
peg #	**RAST Annotation**	**MEROPS/pfam identity**
4	D-alanyl-D-alanine dipeptidase (EC 3.4.13.22)	vanX D-Ala-D-Ala dipeptidase, subfamily M15D unassigned peptidases
163	Zn-dependent peptidase, insulinase family	subfamily M16C unassigned peptidases
173	Membrane proteins related to metalloendopeptidases	subfamily M23B unassigned peptidases
175	Membrane-associated zinc metalloprotease	subfamily M50B unassigned peptidases, RseP peptidase
396	Carboxypeptidase A1 precursor (EC 3.4.17.1)	subfamily M14B unassigned peptidases, CPG70 carboxypeptidase
431	HtrA protease/chaperone protein	subfamily S1C unassigned peptidases, MucD peptidase, Trypsin
464	Proline iminopeptidase (EC 3.4.11.5)	family S33 unassigned peptidases
528	Protease IV	subfamily S49B unassigned peptidases
763	putative metallopeptidase	family M1 unassigned peptidases
831	Lysyl endopeptidase (EC 3.4.21.50)	subfamily S1D unassigned peptidases, pepK
863	Protease II (EC 3.4.21.83)	Prolyl oligopeptidase
890	Peptidase, M23/M37 family	subfamily M23B non-peptidase homologues
922	Putative membrane-bound ClpP-class protease associated with aq_911	NfeD-like C-terminal, partner-binding, subfamily S49C unassigned peptidases
1069	Carboxy-terminal processing protease	CtpC peptidase, CtpA peptidase (Synechocystis-type), C-terminal processing peptidase-3, subfamily S41A unassigned peptidases
1144	Signal peptidase I (EC 3.4.21.89)	subfamily S26A unassigned peptidases
1151	Subtilisin-like serine proteases	subfamily S8A unassigned peptidases (subtilase family)
1229	Probable dipeptidase (EC 3.4.-.-)	family C69 unassigned peptidases
1332	Membrane alanine aminopeptidase N (EC 3.4.11.2)	family M1 unassigned peptidases
1363	Multimodular transpeptidase-transglycosylase (EC 2.4.1.129)	Penicillin binding protein transpeptidase domain,Transglycosylase
1403	Arginine-specific cysteine proteinase	family C25 unassigned peptidases, gingipain K

Table 5, continued.

bin 1		
peg #	**RAST Annotation**	**MEROPS/pfam identity**
1500	Dipeptidyl peptidase IV	subfamily S9B unassigned peptidases
2079	Putative carboxy-terminal processing protease (EC 3.4.21.102)	C-terminal processing peptidase-3, subfamily S41A unassigned peptidases, CtpC peptidase
2088	Peptidase, family M23 (EC 3.4.24.-)	subfamily M23B unassigned peptidases
2102	Prolyl endopeptidase (EC 3.4.21.26)	subfamily S9A unassigned peptidases
2230	CAAX prenyl protease 1, putative	subfamily M48A unassigned peptidases
2307	Peptidase, M23,M37 family	subfamily M23B unassigned peptidases
2331	Probable aminopeptidase (EC 3.4.11.-)	subfamily M28F unassigned peptidases
2645	Dipeptidyl peptidase IV	subfamily S9B unassigned peptidases
2645	Dipeptidyl peptidase IV	subfamily S9B unassigned peptidases
2790	Peptidase, M50 family	family M50 unassigned peptidases
2836	D-alanyl-D-alanine carboxypeptidase	family S12 unassigned peptidases, Beta-lactamase
Total Exported Peptidases		32/69

Table 5, continued.

Bin 2		
Peg #	**RAST Annotation**	**MEROPS Identity**
661	Dipeptidyl peptidase IV	subfamily S9B unassigned peptidases, prolyl tripeptidyl peptidase
678	Isoaspartyl aminopeptidase (EC 3.4.19.5) Asp-X dipeptidase	family T2 unassigned peptidases
321	Metalloprotease MEP2	Outer membrane protein Omp28
325	Protease precursor	subfamily S8A non-peptidase homologues, subfamily S8A unassigned peptidases (subtilase family)
59	Peptidase, M23/M37 family	subfamily M23B unassigned peptidases, Mername-AA292 peptidase
90	Peptidase, M16 family	subfamily M16B unassigned peptidases
398	Peptidase M23B	subfamily M23B unassigned peptidases
400	aminopeptidase	family M1 unassigned peptidases
527	Prolyl endopeptidase (EC 3.4.21.26)	subfamily S9A unassigned peptidases
250	Carboxy-terminal processing protease	CtpC peptidase, CtpA peptidase (Borrelia-type), CtpA peptidase (Synechocystis-type), S41A unassigned peptidases, C-terminal processing peptidase-3, subfamily S41A unassigned peptidases
277	HtrA protease/chaperone protein	CtpC peptidase, CtpA peptidase (Borrelia-type), CtpA peptidase (Synechocystis-type) Bacteroides salanitronis S41A unassigned peptidases, C-terminal processing peptidase-3, subfamily S41A unassigned peptidases
325	Protease precursor	subfamily S8A non-peptidase homologues, subfamily S8A unassigned peptidases (subtilase family)
481	Metalloprotease MEP2	subfamily S8A unassigned peptidases
Total Exported Peptidases		13/20

Table 6. Type IX secretion system (T9SS) and gliding motility genes found in bins 1 and 2.

	bin 1			bin 2		
Gene	**peg**	**RAST annotation**	**e value**	**peg**	**RAST annotation**	**e value**
gldA	779	ABC transporter, ATP-binding protein	1E-107	293	ABC transporter ATP-binding protein	7E-22
gldB	2746	GldB	N/A	-	-	-
gldD	671	GldD	N/A	-	-	-
gldF	1623	gliding motility protein GldF	N/A	331	gliding motility protein GldF	N/A
gldG	1622	gliding motility protein GldG	N/A	332	gliding motility protein GldG	N/A
gldH	1694	GldH		-	-	-
gldI	2830	Probable peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8)	3E-14	429	Probable peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8)	2E-11
gldJ	951	GldJ	N/A	-	-	-
porK/gldK	2778	GldJ	1E-106	-	-	-
porL/gldL	2777	hypothetical protein	3E-30	-	-	-
porM/gldM	2776	hypothetical protein	2E-59	-	-	-
porN/gldN	2775	GldN	6E-16	-	-	-
remA	632	putative autotransporter protein	4E-13	-	-	-
sov/sprA	472	hypothetical protein	0	-	-	-
sprB	327	hypothetical protein	4E-16	636	internalin, putative	2 E-53
porW/sprE	172	TPR domain protein	1E-64	396	TPR domain protein	5E-55
porT/sprT	2364	PorT protein	N/A	-	-	-
porX	853	Response regulator	0	-	-	-
porY	466	sensor histidine kinase	3E-90	386	Nitrogen regulation protein NtrY	5E-18
ompA	2001	outer membrane lipoprotein omp16 precursor	3E-98	317	Outer membrane lipoprotein omp16 precursor	2E-40
porQ	168	low affinity penicillin binding protein	6E-29	-	-	-
porP	2779	hypothetical protein	4E-34	-	-	-
porV	955	hypothetical protein	5E-96	-	-	-

Table 6, continued.

	bin 1			bin 2		
Gene	**peg**	**RAST annotation**	**e value**	**peg**	**RAST annotation**	**e value**
β barrel outer membrane protein	1727	hypothetical protein	7E-11	-	-	-
tonB-dependent receptor; β barrel protein	1708	putative ferric aerobactin receptor	1E-15	81	TonB-dependent receptor, plug precursor	1E-15
omp17	1730	Outer membrane protein H precursor	2E-38	-	-	-
porU	953	hypothetical protein	1E-151	-	-	-
porZ	1417	Immunoreactive 84kD antigen PG93	6E-70	466	Immunoreactive 84kD antigen PG93	3E-42