

# Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics

Cody S. Sheik,<sup>1</sup> Sunit Jain<sup>1</sup> and Gregory J. Dick<sup>1,2,3\*</sup>

<sup>1</sup>Department of Earth and Environmental Sciences and

<sup>2</sup>Center for Computational Medicine and Bioinformatics and

<sup>3</sup>Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA.

## Summary

**Chemolithotrophy is a pervasive metabolic lifestyle for microorganisms in the dark ocean. The SAR324 group of *Deltaproteobacteria* is ubiquitous in the ocean and has been implicated in sulfur oxidation and carbon fixation, but also contains genomic signatures of C<sub>1</sub> utilization and heterotrophy. Here, we reconstructed the metagenome and metatranscriptome of a population of SAR324 from a hydrothermal plume and surrounding waters in the deep Gulf of California to gain insight into the genetic capability and transcriptional dynamics of this enigmatic group. SAR324's metabolism is signified by genes that encode a novel particulate hydrocarbon monooxygenase (pHMO), degradation pathways for corresponding alcohols and short-chain fatty acids, dissimilatory sulfur oxidation, formate dehydrogenase (FDH) and a nitrite reductase (NirK). Transcripts of the pHMO, NirK, FDH and transporters for exogenous carbon and amino acid uptake were highly abundant in plume waters. Sulfur oxidation genes were also abundant in the plume metatranscriptome, indicating SAR324 may also utilize reduced sulfur species in hydrothermal fluids. These results suggest that aspects of SAR324's versatile metabolism (lithotrophy, heterotrophy and alkane oxidation) operate simultaneously, and may explain SAR324's ubiquity in the deep Gulf of California and in the global marine biosphere.**

## Introduction

The deep ocean represents one of the largest yet most understudied microbial biomes on earth (Orcutt *et al.*, 2011). Scarcity of surface-derived labile carbon within the

deep ocean (Aristegui *et al.*, 2009) is thought to constrain heterotrophic microbial activity. However, recent identification of carbon fixation genes in the dark ocean (> 200 m below surface) (Swan *et al.*, 2011) underscores autotrophy as a metabolic strategy for many dark ocean-associated micro-organisms (Karl *et al.*, 1984). Because autotrophy is energetically taxing (Hügler and Sievert, 2011), organisms must couple carbon fixation to an electron-donating reaction such as the oxidation of ammonia, nitrite, sulfur, iron, methane, hydrogen or formate. Yet such electron donors required to power carbon fixation in the deep ocean are scarce (Reinthal and Van, 2010), suggesting that mixotrophy (Sorokin, 2003; Moran *et al.*, 2004; Dick *et al.*, 2008) may be necessary for micro-organisms to persist until preferred energy sources are encountered.

Within the deep ocean, hydrothermal vents are crucial sources of reduced compounds such as sulfur, ammonia, iron, manganese and methane that fuel primary production (Orcutt *et al.*, 2011; Dick *et al.*, 2013). Thus, hydrothermal vents represent oceanic oases for microbial activity and are excellent *in situ* laboratories in which to investigate microbial responses to pulses of electron donors and potentially to disentangle the ecology and physiology of deep-sea micro-organisms. At the hydrothermal vent orifice, emitted hydrothermal fluids meet cold seawater and form buoyant hydrothermal plumes that rise hundreds of meters above the seafloor (Holden *et al.*, 2012). Within these plumes, stable mineral aggregates are formed that contain appreciable fractions of microbial cells and exopolysaccharides (Toner *et al.*, 2009), suggesting micro-organisms actively exploit these microenvironments and potentially influence the speciation of minerals and fate of incorporated organic compounds.

Despite long-standing appreciation that micro-organisms significantly contribute to the geochemistry of hydrothermal plumes (Karl *et al.*, 1980; Jannasch and Mottl, 1985; Winn *et al.*, 1986), plume microbiology has not been studied from a genetic perspective until recently. Molecular characterization of plume communities at Guaymas Basin (GB) in the Gulf of California (Dick and Tebo, 2010) revealed microbial communities shaped by the unusual geochemistry arising from hydrothermal–sediment interactions that result in fluids enriched with

Received 13 August, 2012; revised 9 May, 2013; accepted 24 May, 2013. \*For correspondence. E-mail gdick@umich.edu; Tel. 1 734 763 3228; Fax 1 734 763 4690.

manganese, ammonia, methane and short-chain alkanes but deficient in free sulfide and iron relative to typical eastern Pacific Rise hydrothermal fluids (Von Damm *et al.*, 1985; Welhan and Lupton, 1987). Resultant plume microbial communities are enriched with micro-organisms that perform methanotrophy, sulfur oxidation, ammonia oxidation and heterotrophy (Lesniewski *et al.*, 2012). Interestingly, members of the novel deltaproteobacterial clade SAR324, also known as Marine Group B (Wright *et al.*, 1997), were identified as abundant members of GB hydrothermal plumes (Dick and Tebo, 2010; Lesniewski *et al.*, 2012).

SAR324 is a deep-branching phylogenetic clade within the *Deltaproteobacteria* that was first identified in the Sargasso Sea and northeastern subarctic Pacific Ocean (Wright *et al.*, 1997). SAR324 has since been identified in diverse environments (Dick *et al.*, 2013) such as the North Pacific Subtropical Gyre (DeLong *et al.*, 2006), tropical marine ecosystems (Brown and Donachie, 2007), the Saanich Inlet oxygen minimum zone (OMZ) (Zaikova *et al.*, 2010), the Black Sea redox boundary (Fuchsman *et al.*, 2011), and deep ocean and polar waters (Ghiglione *et al.*, 2012). Analysis of OMZs shows that SAR324 are among the most frequently encountered taxa based on 16S rRNA gene libraries, and the abundance of SAR324 correlates with low-oxygen conditions (Wright *et al.*, 2012).

Despite this ubiquity and abundance, little is known of the physiology of SAR324. Wright and colleagues (1997) initially suggested the SAR324 cluster may participate in the sulfur cycle given their association within the *Deltaproteobacteria*. Fosmids containing SAR324 16S rRNA genes recovered from deep samples at station ALOHA (DeLong *et al.*, 2006) and Monterey Bay (Rich *et al.*, 2011) yielded little insight into the physiology of this cryptic group but suggest the potential for organic phosphate degradation (Brown and Donachie, 2007). Recently, single-cell sequencing produced three SAR324 genomes, two from the mesopelagic (~800 m below surface) that revealed capabilities for sulfur oxidation and metabolic flexibility (Swan *et al.*, 2011), and one from a coastal environment (6 m below surface) that suggests a

heterotrophic lifestyle including degradation of chlorophyll lipids (Chitsaz *et al.*, 2011).

Although tremendous strides have been made in elucidating the potential metabolic versatility of SAR324, the environments and conditions in which these diverse metabolisms are active remain unknown. Here, we present metagenomic and metatranscriptomic results from an abundant population of SAR324 in the deep GB (GB-SAR324) that indicate genes for alkane oxidation, heterotrophy and lithotrophy are all present and actively expressed, supporting the notion that SAR324 utilizes diverse sources of energy and carbon within the same environment to sustain its metabolism.

## Results and discussion

### *Characteristics of the Guaymas Basin SAR324 (GB-SAR324) metagenomic bin*

*De novo* metagenomic assembly was performed using shotgun pyrosequencing reads from four samples, two within the GB hydrothermal plume (1950 and 1963 m), one from just above the plume (1600 m) and one from neighboring Carmen Basin (1900 m), which does not host hydrothermal activity. Genomic binning of contigs (length >2500 bases) with emergent self-organizing maps (ESOM) of tetranucleotide frequency (Dick *et al.*, 2009b) produced a well-defined genomic bin (Fig. S1) that contained two non-overlapping contigs with 16S rRNA genes that were 98–99% similar to previously reported SAR324 sequences (Table 1). This SAR324 genomic bin contains 2.5 Mb of consensus sequence from a total of 512 contigs with average length of 5.4 Kb and average coverage of 6× (Table 1). These contigs contained nearly 3600 predicted genes, including 73% of universally conserved bacterial genes expected to be present (Martin *et al.*, 2006) (Table S1). Comparison of this partial GB-SAR324 genome to two recently published dark ocean-associated (~800 m depth) single-cell amplified partial genomes (SAGs) of SAR324 (Swan *et al.*, 2011) revealed high average nucleotide identity (Fig. S2, AFIA 96% and AFIB 97% average BLASTn similarity). Approximately 55% of genes are shared between GB-SAR324 and dark ocean

**Table 1.** Description of Guaymas Basin SAR324 metagenome and comparison to other SAR324-cluster single amplified genomes and fosmids.

	Guaymas Basin	Single amplified genomes			Fosmids	
		AAA240-J09	AAA001-C10	JCVI	AY458630	GU474888
Size (Mb)	2.5	2.3	2.2	4.3	0.44	0.34
No. contigs	512	273	190	823	1	1
Max contig length (kb)	25.9	106	170	113	44	34
No. genes	3536	2270	2105	3811	43	61
Coverage	6×	–	–	–	–	–
% G + C	42.5	41	41	43	42.6	42.4
16S similarity (%)	–	97.8	99.3	89.3	99.9	99.8

SAR324 SAGs. Dark ocean SAGs also recruited many GB metagenomic pyrosequencing reads (Fig. S3). Comparison to the surface SAR324 SAG (Chitsaz *et al.*, 2011) resulted in few shared genes (< 100 genes) and low sequence similarity (< 90%). Percent G + C content of all SAR324 (SAGs and fosmids) recovered to date is similar to that of GB-SAR324 (42.5) and ranges from 41.1 to 43 for single-cell genomes, while fosmids (AY4586630 and GU474888) are 42.6 and 42.4 (Rich *et al.*, 2011).

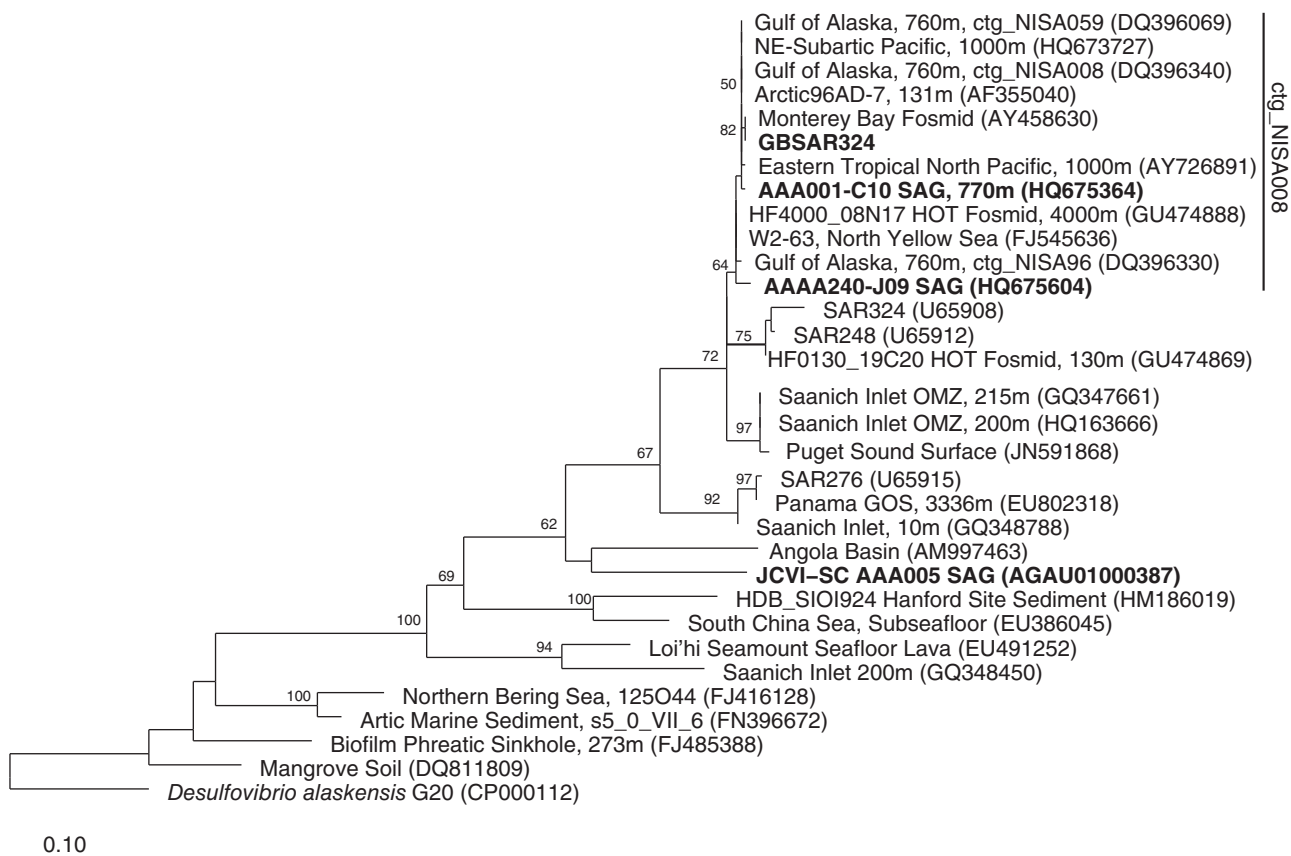
#### Phylogeny and distribution of SAR324 in GB

Phylogenetic analysis showed that the GB-SAR324 16S rRNA gene is affiliated with the ctg\_NISA008 clade and shares 98–100% sequence identity with others in this group (Fig. 1) (Penn *et al.*, 2006; Pham *et al.*, 2008). This clade contains other pelagic SAR324 including two sequenced fosmids (ALOHA and Monterey Bay) and the two dark ocean single amplified genomes (Swan *et al.*, 2011) (Fig. 1; see Table 1 for % similarities of 16S rRNA genes). The GB-SAR324 16S rRNA gene sequence from the metagenome was nearly identical (~ 99% BLASTn) to

sequences from clone libraries generated from the same samples (Dick and Tebo, 2010). Abundance of SAR324 in bacterial clone libraries was 7 and 10.5% of clones within the plume and 5 and 11% in the background (Dick and Tebo, 2010). In 16S rRNA tagged pyrosequencing libraries from samples throughout the full depth profile of the GB water column, total abundance of SAR324 ranged from undetected to nearly 3% of the total community (Fig. S4), and a total of 20 SAR324 phylotypes (operational taxonomic units clustered at 97% similarity) that were 88–100% similar to the metagenomic 16S rRNA sequence were identified. However, the lack of sequence variation in reads contributing to metagenomic contigs suggests that the metagenomic and metatranscriptomic data presented here are derived primarily from a single dominant population of SAR324.

#### Central carbon metabolic pathways suggest a mixotrophic lifestyle of GB-SAR324

Reconstruction of central metabolic pathways in GB-SAR324 revealed most of genes for the oxidative



**Fig. 1.** Phylogenetic relationship of Guaymas Basin SAR324 to other Marine Group B/SAR324-cluster SSU rRNA sequences identified in the Silva v. 111 database. Guaymas Basin and single amplified genome sequences are in bold.

tricarboxylic acid (TCA) cycle, glycolysis/gluconeogenesis and the Calvin–Benson–Bassham (CBB) pathways (Table S2). Also identified were genes encoding long- and short-chain fatty acid dehydrogenases, zinc-containing and non-specific alcohol dehydrogenases, aromatic dioxygenases, a lactate/malate dehydrogenase and most of the  $\beta$ -oxidation pathway for breakdown of fatty acids (Tables S2 and S3). A gene for incorporation of odd chain fatty acids (from  $\beta$ -oxidation or propionate) into the TCA cycle was present [propionyl-coenzyme A (CoA) carboxylase], suggesting that bicarbonate is incorporated during the conversion of propionyl-CoA to methylmalonyl-CoA. Such anapleurotic incorporation of bicarbonate during heterotrophic breakdown of fatty acids could contribute to uptake of labelled bicarbonate by particle-associated SAR324 observed by Swan and colleagues (2011).

Like mesopelagic SAR324 SAGs, the potential for carbon fixation through the CBB pathway is present in GB-SAR324 as a gene encoding the large subunit of ribulose biphosphate-1,5-carboxylase oxygenase (RuBisCO) was identified. This *cbbL* gene is phylogenetically associated with form II and is most closely related to genes from single-cell amplified SAR324 genomes and the cultured sulfate-reducing deltaproteobacterium *Desulfovibro aespoeensis* Aspo-2 (Fig. S5). Although form I RuBisCO is present in other SAR324 genomes (Swan *et al.*, 2011), it was not detected in GB-SAR324 (Fig. S5), suggesting that SAR324 may contain either or both forms of RuBisCO. Interestingly, form II *cbbL* is generally thought to be adapted to low oxygen, high CO<sub>2</sub> environments (Badger and Bek, 2008) and is also present and active in SUP05 populations that are also abundant in the deep GB (Anantharaman *et al.*, 2013). Thus, this form II RuBisCO seems to be well adapted to the low-O<sub>2</sub> conditions present in both the deep portions of GB and potentially the upper water column OMZ (400–1000 m) (Fig. S4).

#### Evidence for hydrocarbon and C<sub>1</sub> metabolism

Methanotrophy in SAR324 was inferred by Swan and colleagues (2011) based on the presence of genes similar to particulate methane monooxygenase (pMMO) and other C<sub>1</sub> carbon metabolism genes. Although no pMMO-like genes were identified in the initial GB-SAR324 bin (contigs > 2500 bp), shorter metagenomic and metatranscriptomic sequences with high similarity to the SAR324-SAG pMMO genes were identified (Figs S6 and S7). Phylogenetic analysis showed the *pmmoA*-like gene is affiliated with sequences from organisms known to oxidize C<sub>2</sub>-C<sub>4</sub> alkanes and should be classified as a particulate hydrocarbon monooxygenase (pHMO) (Fig. 2; Coleman *et al.*, 2011;

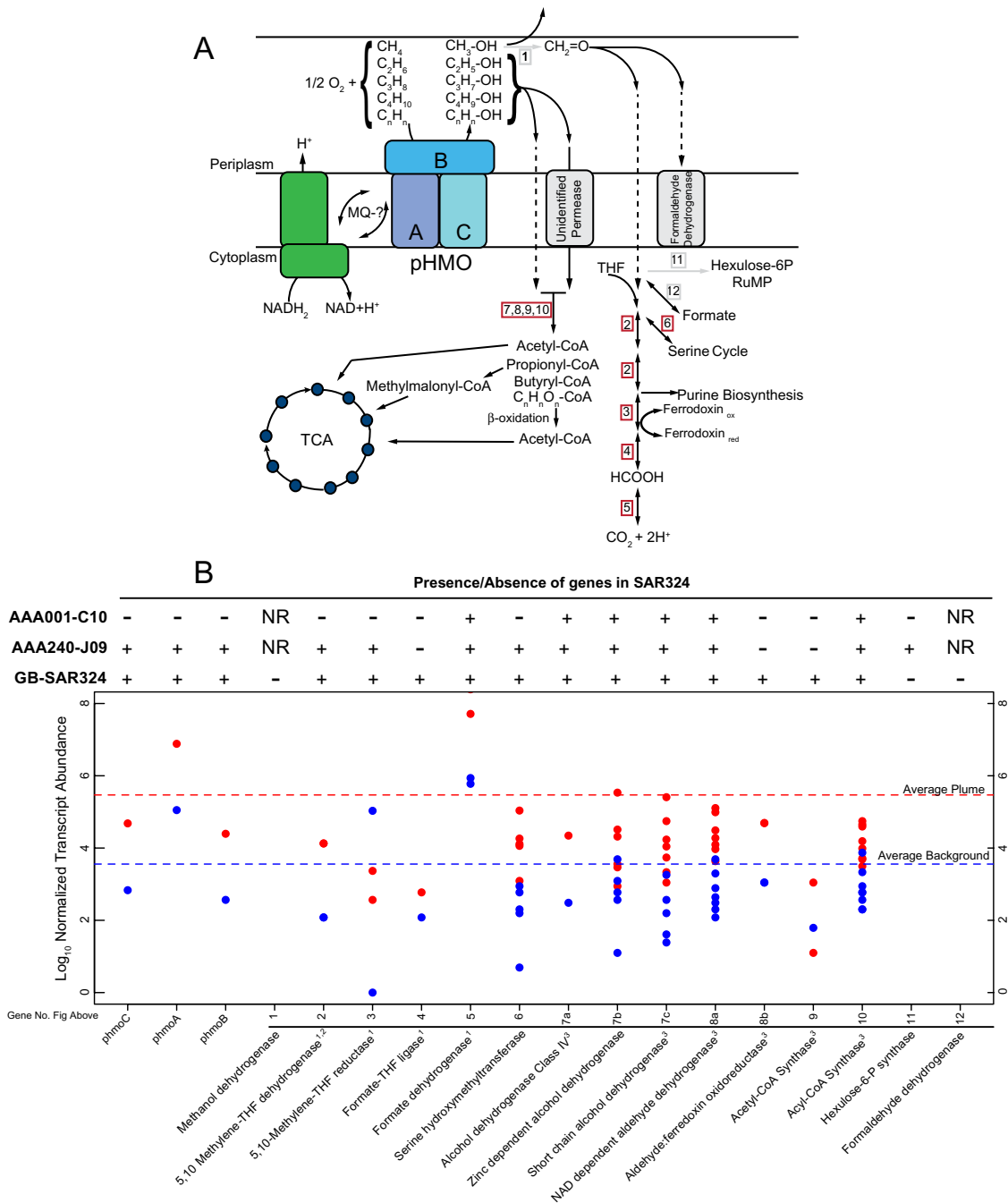
Sayavedra-Soto *et al.*, 2011; see Li *et al.*, 2013 for an in depth analysis of overall pMMO diversity, abundance and distribution at GB).

An array of genes for the utilization of oxidized hydrocarbon breakdown products was identified in the GB-SAR324 metagenomic bin (Fig. 3). Several short- to mid-chain primary alcohol dehydrogenases are present, consistent with the conversion of hydrocarbon-derived alcohols to aldehydes for downstream degradation. Oxidation of the acyl-CoA derivatives could occur through several routes (Fig. 3A) depending on carbon chain length. Genomic evidence for secondary alcohol utilization was also identified by the presence of a 3-hydroxyisobutyrate dehydrogenase/ beta-hydroxyacid dehydrogenase. Most of these genes are also present in both SAR324 mesopelagic SAGs (Fig. 3B).

While the direct utilization of methane cannot be confirmed solely through genome analysis, several genomic signatures of methylotrophy are present. Genes encoding a periplasmic methanol dehydrogenase were not identified in GB-SAR324, and none of the identified alcohol dehydrogenases contains twin-arginine translocation signal peptides for periplasmic excretion. Although this suggests that GB-SAR324 may not utilize methanol via a traditional methylotroph methanol dehydrogenase, the presence of NAD(P)-binding alcohol dehydrogenases in GB-SAR324 suggests alternative routes for methanol conversion (Chistoserdova, 2011), thus methanol utilization cannot be fully ruled out. In addition, genes necessary for formaldehyde detoxification or incorporation of downstream products of methane oxidation via tetrahydrofolate intermediates and portions of the serine pathway were present including a putative cytoplasmic NAD(P)-binding formaldehyde dehydrogenase (Table S2). Finally, a putative dimethylsulfoniopropionate (DMSP) demethylase was identified and suggests that this SAR324 group could utilize portions of the C<sub>1</sub> pathway for DMSP degradation (Chistoserdova, 2011).

A periplasmic oriented, membrane-associated formate dehydrogenase (FDH) is also present and potentially linked to the quinone pool through a membrane spanning NrfD (Pereira *et al.*, 2011), which is adjacent to the FDH on contig 151. FDHs are also in both SAR324 SAGs and are highly similar to GB-SAR324 (97–98% nucleotide similarity). No hydrogenases could be identified. Also present was an aerobic carbon monoxide dehydrogenase. Carbon monoxide oxidation would provide GB-SAR324 cells with an additional source of energy especially within hydrothermal vent fluids (Lilley *et al.*, 1982) and could serve as a hydrogen source for other micro-organisms when SAR324 are attached to particles, a lifestyle suggested previously (Fuchsman *et al.*, 2011; Swan *et al.*, 2011).





**Fig. 3.** Depiction of potential hydrocarbon degradation routes based on annotated genes from the Guaymas Basin SAR324 genome bin. A. Potential routes for hydrocarbon oxidation; numbers 1–12 correspond to enzymes shown in B. B. Presence of hydrocarbon oxidation genes in dark ocean SAR324 single amplified genomes and GB-SAR324 (top) and the transcriptional response of these genes identified at Guaymas Basin within the plume (red) and background (blue) samples. Genes not identified in GB-SAR324 are depicted in grey. <sup>1</sup>As reported by Swan and colleagues (2011). <sup>2</sup>In GB-SAR324 these genes are predicted to be bi-functional proteins similar to *foID* in *Clostridia* species. <sup>3</sup>Determined using BLASTn of GB-SAR324 contigs against contigs from both SAR324 SAGs. NR = not reported, THF = tetrahydrofolate.

### Terminal electron acceptors and energy harvesting complexes

Reconstruction of electron transport chains in GB-SAR324 suggests the potential for aerobic respiration and dissimilatory nitrite reduction. Genes for complex I (NADH: ubiquinone oxidoreductase), complex II (succinate dehydrogenase), many subunits for a putative complex III (quinone: cytochrome *c* oxidoreductase) and high-affinity cytochrome *c* oxidase (complex IV) (Table S3) were identified. No low-affinity cytochrome *c* oxidase was identified, supporting the hypothesis that SAR324 is adapted to low-oxygen concentrations.

A nitrite reductase (NirK), copper-containing form, was identified in both SAR324 SAGs (Swan *et al.*, 2011). Much like the pHMO, no *nirK* was identified in the initial GB-SAR324 bin, but it is present on several smaller contigs (< 2500 bp) with high similarity (98%) to *nirK* from SAR324 SAGs. While no nitrate or nitrous oxide reductases were annotated in the GB-SAR324 metagenomic bin, a polyferredoxin gene that could serve as a *NapH* subunit of periplasmic nitrate reductase was identified. Some sulfur-reducing *Deltaproteobacteria* are known to use NirKs for detoxification of nitrite (Greene *et al.*, 2003). However, the potential for GB-SAR324 to use nitrate and/or nitrite as a terminal electron accepting process is intriguing in that nitrite reduction could theoretically drive both alkane and sulfur oxidation, which has been suggested for sulfur-oxidizing SUP05 *Gammaproteobacteria* (Walsh *et al.*, 2009). Furthermore, nitrite is likely available in GB by way of abundant and transcriptionally active ammonia-oxidizing archaea (Baker *et al.*, 2012; Lesniewski *et al.*, 2012).

### Sulfur cycling

Sulfur oxidation was recently suggested to be a key physiological feature of SAR324 based on the presence of adenosine 5'-phosphosulfate reductase (*aprA*) and reverse-type dissimilatory sulfite reductase (*rdsrA*) in SAGs (Swan *et al.*, 2011). The GB-SAR324 genomic bin also contains genes encoding the full intracellular pathway that is utilized in sulfur oxidation including dissimilatory sulfur oxidation (*dsrAB*), *aprAB*, DSR electron-shuttling genes (*dsrNCDLEFHMNOPJ*), sulfate adenyltransferase (*Sat*) and a quinone-interacting membrane-bound oxidoreductase (*QMO*) complex (Fig. 4, Table S3 and Supporting Information). To our knowledge, this is the first reported occurrence of *dsrEFH* in a *deltaproteobacterium*. These genes co-occur on three primary contigs (41, 352 and 669) that have partial synteny to SAR324 SAGs (Fig. S8).

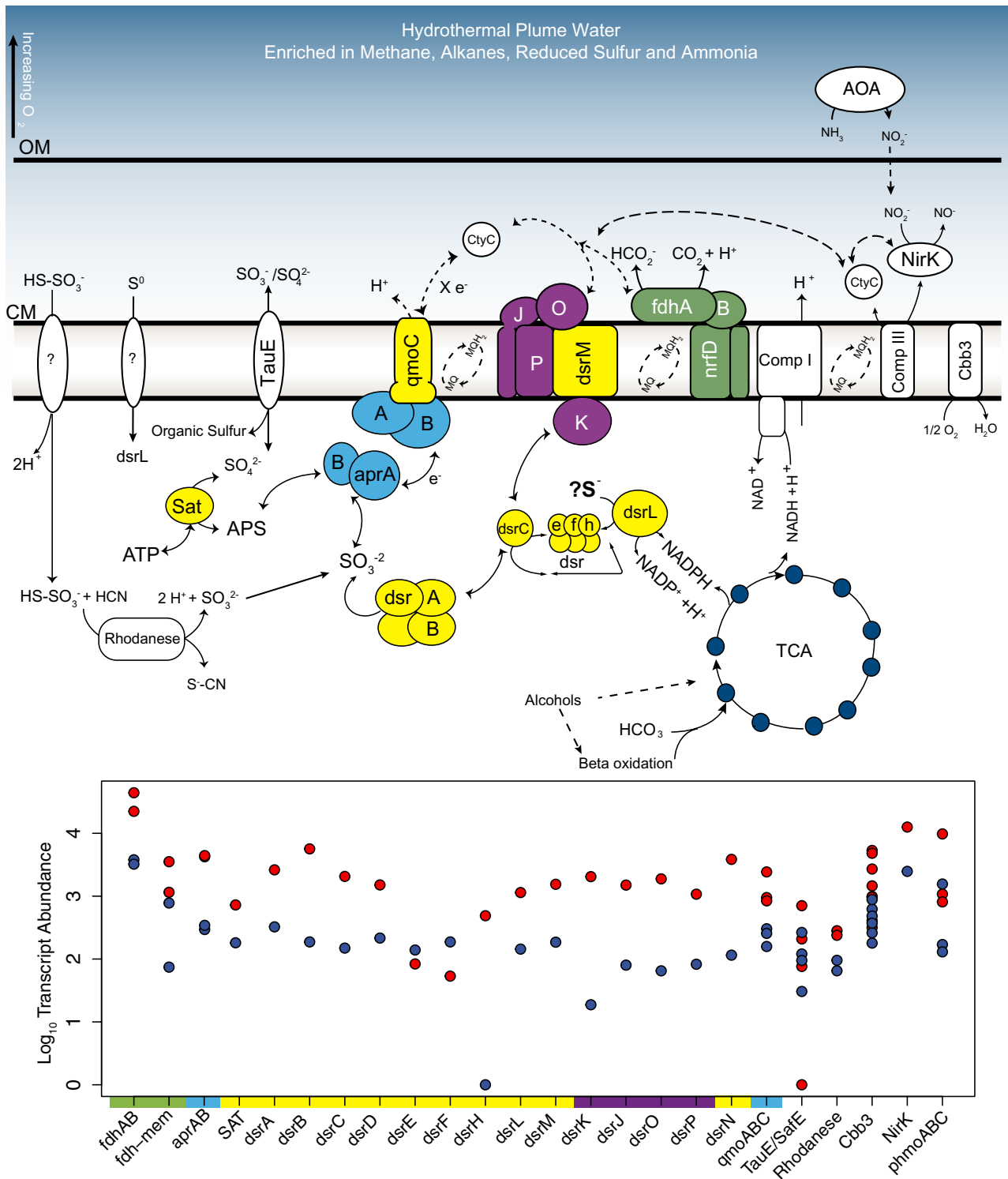
The phylogeny of *dsrAB* and *aprA* are commonly used to infer oxidative versus reductive direction of the path-

way (Meyer and Kuever, 2007; Loy *et al.*, 2009; Swan *et al.*, 2011). The GB-SAR324 *dsrA* gene clustered phylogenetically in a bootstrap supported clade that is distinct from traditional sulfur-oxidizing and sulfur-reducing types (Fig. S9). SAR324 *dsrA* genes are most closely related to those from *Chlorobi* and *Magnetococcus*, but this larger cluster was not bootstrap supported. Similarly, the *aprA* gene forms a larger bootstrap supported clade that includes *Chlorobi* and the sulfur-reducing *Desulfobacca* and *Thermodesulfobivrio* (Fig. S10). The phylogeny of the *Sat* gene, much like *aprA*, places GB-SAR324 in a unique branch associated with but distinctly different from sulfur-oxidizing *Gammaproteobacteria* and two sulfate-reducing bacteria (Fig. S11). Interestingly, the *Chlorobi Sat* genes are not associated with this clade. Phylogeny of these key sulfur utilization genes point to a complicated evolutionary history that is likely punctuated by several horizontal gene transfer events. To date, the balance of evidence suggests these genes are involved in sulfur oxidation via the reverse dissimilatory sulfite reductase (rDSR) pathway.

Given that the classical rDSR system oxidizes sulfide remobilized from sulfur globules that are produced via the Sox system (Pott and Dahl, 1998; Hensen *et al.*, 2006), we searched the GB-SAR324 contigs for sulfide-quinone oxidoreductase and sox genes that mediate periplasmic sulfur oxidation. No sulfide-quinone oxidoreductase or flavocytochrome-C sulfide dehydrogenase was identified. A single gene annotated as *soxZ* was identified; however, the surrounding genes displayed only low sequence similarity to known sox genes. The *soxZ* gene was highly similar to SAR324 SAG AAA001-C10 (99%) but was phylogenetically divergent from known sulfur oxidizers (Fig. S12). Based on the novelty of this sox-like gene, it is difficult to assess SAR324's potential for periplasmic sulfur oxidation.

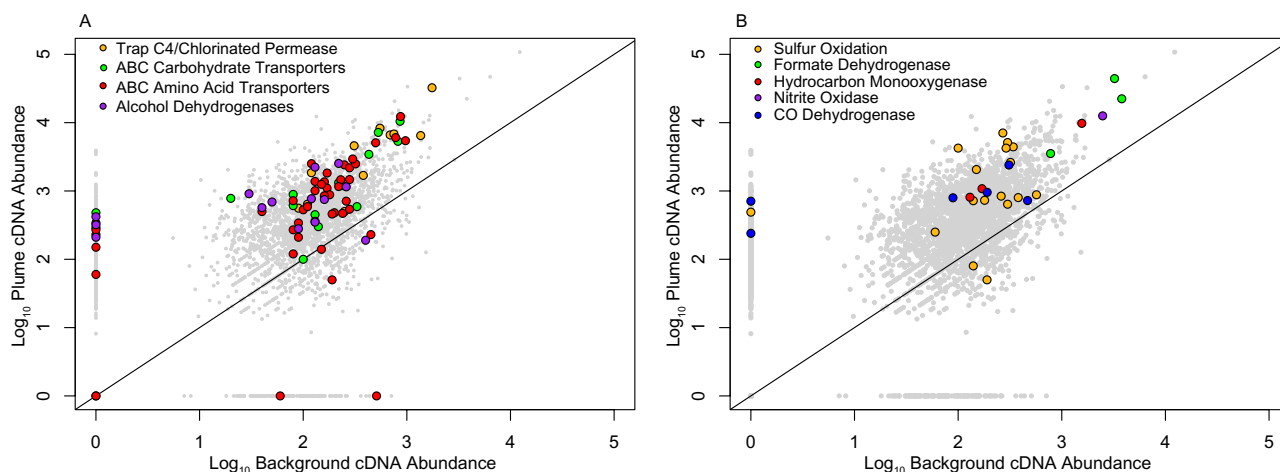
### Metatranscriptomic response of GB-SAR324 to hydrothermal plumes

We utilized a shotgun metatranscriptomic approach to assess gene expression by GB-SAR324 in the plume and background environments. While transcript abundance does not necessarily correlate directly to cellular function or process rate, metatranscriptomics provides valuable insights into microbial responses to changing environmental conditions (Moran *et al.*, 2013). Here, we used the plume as an *in situ* laboratory to measure the transcriptomic response of GB-SAR324 to sulfur and hydrocarbon-rich conditions. Most GB-SAR324 genes show higher transcript abundance in plume (2784 genes) than background (566 genes) (Fig. 5). However, the number of genes for which transcripts were only detected



**Fig. 4.** Putative intracellular sulfur metabolism in Guaymas Basin SAR324. Depiction of sulfur metabolisms based on annotated genes recovered from the SAR324 ESOM bin. Function of genes in GB-SAR324 is tentative and based upon prior proposed functional models taken from Stockdreher and colleagues (2012), Pereira and colleagues (2011), and Mussmann and colleagues (2005). Colour of genes and proteins denote the contig origin: yellow c41, green c151, purple c352 and blue c669.





**Fig. 5.** Transcriptional abundance of Guaymas Basin SAR324 genes in plume and background. Transcripts were mapped to the GB-SAR324 ESOM bin from Illumina paired end sequencing libraries. A. Highlights carbon and nitrogen uptake and utilization genes. B. Highlights sulfur oxidation respiration, formate dehydrogenases and the pHMO genes.

in background (499) was much higher than genes for which transcripts were only detected in the plume (138), hinting at specialized functions for low nutrient/energy conditions. A majority of those genes with transcripts only detected in background were related to exogenous carbon and nitrogen metabolisms (Fig. S13). Additionally, transcripts indicative of organic sulfur metabolism, specifically alkanesulfonate assimilation and degradation, were only detected in the background, suggesting GB-SAR324 may scavenge sulfonated organic compounds for both carbon and reduced sulfur. Of the genes detected only in the plume metatranscriptome were genes for nitrate/nitrite transport, protein metabolism, oxidative stress response and genes in the serine cycle (Fig. S14).

Among the most abundant genes detected in both plume and background metatranscriptomes were those involved in energy harvesting and carbon metabolism, including FDH, *nirK*, pHMO subunit A (*pHMOA*), two hypothetical proteins, and predicted transporters for carbohydrates, amino acids and four-carbon dicarboxylates (Fig. 5). Nearly all genes involved in hydrocarbon oxidation and subsequent degradation recruited more transcripts in plume than background (Fig. 3), consistent with a transcriptomic response to elevated hydrocarbon concentrations in the plume. However, transcripts for hydrocarbon utilization were also recovered from background, suggesting either that these genes are transcribed constitutively or that GB-SAR324 utilizes hydrocarbons at lower (background) concentrations. While no methanol dehydrogenase was identified, the transcription of pathways necessary for the conversion of formaldehyde to either formate or biomass were expressed in both plume and background at low abundance. Transcripts from tetrahydrofolate ( $H_4F$ ) genes were generally more abun-

dant in the plume than in the background (Table S2). Expression of GB-SAR324's  $H_4F$  pathway suggests that formaldehyde can either proceed to biomass via serine hydroxymethyltransferase or to generate adenosine triphosphate (ATP) from the formate-tetrahydrofolate ligase and subsequently FDH. Interestingly, GB-SAR324 may not utilize a full serine cycle, as many of the genes necessary to recycle glyoxylate back to glycine are not present. Instead, glycine may be recycled from glycolate produced from oxygen reacting with the oxygen sensitive type II RuBisCO.

While the FDH is used in methylotrophy (Chistoserdova, 2011), the abundance of FDH transcripts in the plume and especially background, where methane concentrations are presumably very low, suggests that formate may be an appreciable source of energy that has been overlooked in the deep ocean. Furthermore, presence of a similar FDH in the surface SAR324 SAG suggests that the use of formate may be widely distributed among the SAR324. Formate has a similar reduction potential to hydrogen [Hydrogen:  $E' = -414$  mV, formate:  $-432$  mV; at pH = 7, 25°C, 1 ATM (Thauer *et al.*, 1977)], thus it is an effective electron donor and could not only be coupled to reduction of oxygen or nitrate/nitrite but could replace NADH as the source of electrons to drive the oxidation of hydrocarbons. To our knowledge, no measurement of formate concentration in GB plumes is available; however, formate is enriched in GB sediment hydrothermal fluids via pyrolysis of organic matter in sediments (Martens, 1990). Formate is also produced abiotically at ultramafic hydrothermal vents (Lang *et al.*, 2010), thus it could be an even more important electron donor at ultramafic sites such as the Mid-Cayman Rise where SAR324 is also abundant in plumes (German

*et al.*, 2010). While we cannot exclude intracellular recycling of formate from formaldehyde via methylotrophy, exogenous sources of formate may still be important for supporting growth of methylotrophs (Chistoserdova *et al.*, 2004) and SAR324 in the oceans.

Sulfur oxidation genes were also highly represented in the metatranscriptome, and 16 of the 18 genes for sulfur oxidation recruited more transcripts in plume than background (Figs 4 and 5). Although there is no data available on the concentration or speciation of sulfur in the GB plume, thermodynamic modelling of the GB plume indicates that free sulfide concentrations are low and that elemental or thiosulfate sulfur oxidation coupled to either nitrate or oxygen are the most favourable sulfur-based energy sources (Anantharaman *et al.*, 2013). Indeed, the expression of rhodanese supports the utilization of alternative sulfur species such as thiosulfate (Fig. 4B), and lower abundance of *dsrEFH* transcripts in the plume suggests that free sulfide may not be the primary source of reduced sulfur for GB-SAR324, as *dsrEFH* expression is correlated with sulfide concentration (Falkenby *et al.*, 2011).

The overall picture that emerges from the GB-SAR324 metatranscriptome is that genes encoding diverse metabolic processes including hydrocarbon and sulfur oxidation, heterotrophy, and utilization of both oxygen and nitrite/nitrate as electron acceptors are all actively transcribed in the deep GB. The population-level data we present here cannot distinguish whether different genes are being simultaneously transcribed within the same cell, or whether individual cells specialize in a single process and the pattern we observe stems from population-level heterogeneity. One appealing scenario that supports the former is lithoheterotrophy, in which lithotrophy provides supplemental energy and increases heterotrophic efficiency (Sorokin, 2003; Moran *et al.*, 2004; Dick *et al.*, 2008). In the case of GB-SAR324, ATP and/or proton motive force generated by the oxidation of electron donors (sulfur, formate or hydrocarbons) may support high-affinity acquisition of exogenous organic carbon via ATP-dependent ABC-type transporters (amino acids and carbohydrates) or proton gradient-dependent C4-TRAP transporters (C4 dicarboxylates), all of which are highly expressed.

## Conclusions

This study provides insights into a population of SAR324 that dominates a deep-sea, hydrocarbon rich, low-oxygen setting, which contrasts the suboxic to oxic mesopelagic (AAA240-J09 and AAA001-C10, respectively) and epipelagic (JCVI) environments from which SAR324 genomes have previously been retrieved. Nonetheless, metagenomic and metatranscriptomic evidence pre-

sented here point to conserved themes of SAR324 genomes and confirm its metabolically flexible nature. GB-SAR324 is highlighted by the presence and expression of genes indicative of short-chain alkane and sulfur oxidation, terminal respiratory complexes for nitrite and oxygen, as well as acquisition of exogenous organic carbon and nitrogen, suggesting a facultative aerobic/anaerobic, litho-/organo-, auto/hetero-trophic lifestyle. While its hydrocarbon substrate specificity remains uncertain, there are multiple lines of evidence that the ctg\_NISA008 clade likely oxidizes a range of larger chain alkanes. The putative ability to utilize several electron donors (sulfur, hydrocarbons, C<sub>1</sub> compounds, organic carbon) and acceptors (nitrite or oxygen) emphasizes SAR324's importance in the cycling of carbon, nitrogen and sulfur in the ocean and likely contributes to the ubiquity of SAR324 in marine environments.

## Experimental procedures

Samples were collected aboard the R/V New Horizon as described previously (Dick *et al.*, 2009a; Dick and Tebo, 2010). Briefly, water samples were collected via a CTD-Rosette in 10 l Niskin bottles. On board, water was pressured filtered with N<sub>2</sub> gas onto 142 mm polycarbonate filters, immediately preserved with RNALater (Ambion, Grand Island, NY, USA) and stored at -80°C as instructed by the manufacturer. Total community DNA was extracted as described by Dick and Tebo (2010) from two plume (1775 and 1996 m) and two background (1600 and 1900 m) samples. Total extracted DNA was shotgun sequenced with 454-Titanium chemistry (Lesniewski *et al.*, 2012). *De novo* assembly of all metagenomic reads was performed with Mira (ver. 3.0; Chevreaux *et al.*, 1999) and is described by Lesniewski and colleagues (2012). Assembled contigs were submitted to Integrated Microbial Genomics (IMG Taxon Object ID no. 2061766003, <http://img.jgi.doe.gov>) for gene calling and annotation. SAR324-related sequences were identified using tetranucleotide ESOM binning as described by Dick and colleagues (2009b). Total community RNA was extracted from two plume and two background filters using the mirVana miRNA isolation kit (Ambion), treated with DNAase I and concentrated with the RNeasy MinElute kit (Qiagen, Valencia, CA, USA). cDNA preparation using random amplification was performed as per Shi and colleagues (2009) and Frias-Lopez and colleagues (2008). Total community cDNA was shotgun sequenced with 454 titanium pyrosequencing. Pyrosequencing cDNA reads from the plume (896 665 and 664 240 reads), and background (514 607 and 504 086 reads) were mapped to the GB-SAR324 genomic bin with BWA-SW (ver. 0.6.0) (Li and Durbin, 2010) using default parameters. Two cDNA samples used for pyrosequencing, one plume (1950 m) and one background (1600 m) were selected for 100 base paired end shotgun sequencing by Illumina HiSeq2000 technology at the University of Michigan DNA sequencing core. Illumina sequencing yielded approximately 206 and 245 million cDNA reads (Baker *et al.*, 2013), plume and background respectively, which were quality trimmed using Sickle ([Published 2013. This article is a U.S. Government work and is in the public domain in the USA, \*Environmental Microbiology\*, 16, 304–317](https://github.com/ucdavis-</a></p>
</div>
<div data-bbox=)

bioinformatics/sickle.git) then mapped to the SAR324 metagenome bin with BWA (ver. 0.6.0) (Li and Durbin, 2009) using default parameters. Read abundance was extracted from resulting SAM mapping files with SAMtools (ver. 0.1.18) (Li *et al.*, 2009), and read mapping accuracy was visualized with Integrative Genomics Viewer (Thorvaldsdóttir *et al.*, 2013). The relative abundance of each gene in the transcriptome was first normalized to the length of the gene (cDNA reads/length of gene) then normalized to the entire number of cDNA reads recovered from the total community. Minor differences between pyrosequencing and Illumina sequencing technologies were apparent (Fig. S3) and likely stem from depth of sequencing. Illumina rank abundance profiles were more uniform for both highly abundant genes and especially for rare genes. Thus, all transcriptome analyses were based on the Illumina dataset.

16S rRNA gene phylogeny was assessed by aligning the full-length GB-SAR324 SSU ribosomal gene sequence to the pre-aligned Silva SSU ribosomal database (ver. 111 non-redundant) (Pruesse *et al.*, 2007) implemented in the ARB software package (Ludwig *et al.*, 2004). The GB-SAR324 16S rRNA gene alignment was curated by hand to insure accurate base alignment, then alignment columns were positional filtered and exported from ARB. Ribosomal phylogeny, using a subset of SAR324 representative sequences taken from the Silva database version 111, was reconstructed with RAxML (v. 7.2.8) (Stamatakis, 2006; Stamatakis *et al.*, 2008) with implementation of rapid bootstrapping and the GTRGAMMA model. Protein phylogeny was assessed similarly for APS, RuBisCO, putative *soxZ* and *dsrA*. Representative amino acid sequences were downloaded from GenBank and aligned with Muscle (ver. 3.8.31) (Edgar, 2004) using default parameters. Resulting alignments were filtered using Gblocks (ver. 0.91b) and phylogeny reconstructed with RAxML (ver. 7.2.8) using PROTGAMMAGTR model and rapid bootstrapping parameters. The *dsrA* gene from GB-SAR324 and SAR324 SAGs were imported into ARB and aligned to the pre-aligned *dsrAB* database containing genes from cultured sulfur-oxidizing and sulfur-reducing micro-organisms (Fig. S6) (Loy *et al.*, 2009). Bootstrapped maximum likelihood trees were generated *dsrAB* using the pipeline above. Sequences for APS, *dsrAB* and 16S rRNA genes can be obtained from GenBank accessions JX406430-33 respectively.

### Acknowledgements

This research is funded by the Gordon and Betty Moore Foundation through Grant GBMF2609 to Dr Gregory Dick and by the National Science Foundation (OCE 1029242). We thank Karthik Anantharaman for RNA extraction and cDNA library prep and members of the Dick lab and Dr. Jessica Sieber for their helpful comments, constructive criticism and revisions to the manuscript.

### References

Anantharaman, K., Breier, J.A., Sheik, C.S., and Dick, G.J. (2013) Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proc Natl Acad Sci USA* **110**: 330–335.

- Arístegui, J., Gasol, J.M., Duarte, C.M., and Herndl, G.J. (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* **54**: 1501–1529.
- Badger, M.R., and Bek, E.J. (2008) Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO<sub>2</sub> acquisition by the CBB cycle. *J Exp Bot* **59**: 1525–1541.
- Baker, B.A., Lesniewski, R.A., and Dick, G.J. (2012) Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J* **6**: 2269–2279.
- Baker, B.J., Sheik, C.S., Taylor, C.A., Jain, S., Bhasi, A., Cavalcoli, J.D., and Dick, G.J. (2013) Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling. *ISME J* 1–12. doi: 10.1038/ismej.2013.85
- Brown, M.V., and Donachie, S.P. (2007) Evidence for tropical endemicity in the Deltaproteobacteria marine group B/SAR324 bacterioplankton clade. *Aquat Microb Ecol* **46**: 107–115.
- Chevreur, B., Wetter, T., and Suhai, S. (1999) Genome sequence assembly using trace signals and additional sequence information. German Conference on Bioinformatics GCB'99 GCB. *Comput Sci Biol Proc* 45–56.
- Chistoserdova, L. (2011) Modularity of methylotrophy, revisited. *Environ Microbiol* **13**: 2603–2622.
- Chistoserdova, L., Laukel, M., Portais, J., Vorholt, J.A., and Lindstrom, M.E. (2004) Multiple formate dehydrogenase enzymes in the facultative methylotrophy *Methylobacterium extorquens* AM1 are dispensable for growth on methanol. *J Bacteriol* **186**: 22–28.
- Chitsaz, H., Yee-Greenbaum, J.L., Tesler, G., Lombardo, M.J., Dupont, C.L., Badger, J.H., *et al.* (2011) Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nat Biotechnol* **29**: 915–921.
- Coleman, N.V., Yau, S., Wilson, N.L., Nolan, L.M., Migocki, M.D., Ly, M., *et al.* (2011) Untangling the multiple monooxygenases of *Mycobacterium chubuense* strain NBB4, a versatile hydrocarbon degrader. *Environ Microbiol Rep* **3**: 297–307.
- DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.-U., *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- Dick, G.J., and Tebo, B.M. (2010) Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ Microbiol* **12**: 1334–1347.
- Dick, G.J., Podell, S., Johnson, H.A., Rivera-Espinoza, Y., Bernier-Latmani, R., McCarthy, J.K., *et al.* (2008) Genomic insights into Mn(II) oxidation by the marine alphaproteobacterium *Aurantimonas* sp. strain SI85-9A1. *Appl Environ Microb* **74**: 2646–2658.
- Dick, G.J., Clement, B.G., Webb, S.M., Fodrie, F.J., Bargar, J.R., and Tebo, B.M. (2009a) Enzymatic microbial Mn(II) oxidation and Mn biooxide production in the Guaymas Basin deep-sea hydrothermal plume. *Geochim Cosmochim Acta* **73**: 6517–6530.
- Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., and Banfield, J.F. (2009b) Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.

- Dick, G.J., Anantharaman, K., Baker, B.J., Li, M., Reed, D.C., and Sheik, C.S. (2013) The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. *Front Ext Microbiol* **4**: 1–16.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Falkenby, L., Szymanska, M., Holkenbrink, C., Habicht, K., Andersen, J., Miller, M., and Frigaard, N.-U. (2011) Quantitative proteomics of *Chlorobaculum tepidum*: insights into the sulfur metabolism of a phototrophic green sulfur bacterium. *FEMS Microbiol Lett* **323**: 142–150.
- Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and DeLong, E.F. (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105**: 3805–3810.
- Fuchsman, C.A., Kirkpatrick, J.B., Brazelton, W.J., Murray, J.W., and Staley, J.T. (2011) Metabolic strategies of free-living and aggregate-associated bacterial communities inferred from biologic and chemical profiles in the Black Sea suboxic zone. *FEMS Microbiol Ecol* **78**: 586–603.
- German, C., Bowen, A., Coleman, M., Honig, D., Huber, J., Jakuba, M., *et al.* (2010) Diverse styles of submarine venting on the ultraslow spreading Mid-Cayman Rise. *Proc Natl Acad Sci USA* **107**: 14020–14025.
- Ghiglione, J.-F., Galand, P., Pommier, T., Pedrós-Alió, C., Maas, E., Bakker, K., *et al.* (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc Natl Acad Sci USA* **109**: 17633–17638.
- Greene, E., Hubert, C., Nemat, M., Jenneman, G., and Voordouw, G. (2003) Nitrite reductase activity of sulphate-reducing bacteria prevents their inhibition by nitrate-reducing, sulphide-oxidizing bacteria. *Environ Microbiol* **5**: 607–617.
- Hensen, D., Sperling, D., Trüper, H.G., Brune, D.C., and Dahl, C. (2006) Thiosulphate oxidation in the phototrophic sulphur bacterium *Allochromatium vinosum*. *Mol Microbiol* **62**: 794–810.
- Holden, J., Breier, J., Rogers, K., Schulte, M., and Toner, B. (2012) Biogeochemical processes at hydrothermal vents: microbes and minerals, bioenergetics, and carbon fluxes. *Oceanography* **25**: 196–208.
- Hügler, M., and Sievert, S.M. (2011) Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Annu Rev Mar Sci* **3**: 261–289.
- Jannasch, H.W., and Mottl, M.J. (1985) Geomicrobiology of deep-sea hydrothermal vents. *Science* **229**: 717–725.
- Karl, D.M., Wirsén, C.O., and Jannasch, H.W. (1980) Deep-sea primary production at the Galapagos hydrothermal vents. *Science* **207**: 1345–1347.
- Karl, D.M., Knauer, G.A., Martin, J.H., and Ward, B.B. (1984) Bacterial chemolithotrophy in the ocean is associated with sinking particles. *Nature* **309**: 54–56.
- Lang, S.Q., Butterfield, D.A., Schulte, M., Kelley, D.S., and Lilley, M.D. (2010) Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City hydrothermal field. *Geochim Cosmochim Acta* **74**: 941–952.
- Lesniewski, R.A., Jain, S., Anantharaman, K., Schloss, P.D., and Dick, G.J. (2012) The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J* **6**: 1–11.
- Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Li, H., and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**: 589–595.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Li, M., Jain, S., Baker, B.J., Taylor, C., and Dick, G.J. (2013) Novel hydrocarbon monooxygenase genes in the metatranscriptome of a natural deep-sea hydrocarbon plume. *Environ Microbiol*. doi: 10.1111/1462-2920.12182.
- Lilley, M.D., de Angelis, M.A., and Gordon, L.I. (1982) CH<sub>4</sub>, H<sub>2</sub>, CO and N<sub>2</sub>O in submarine hydrothermal vent waters. *Nature* **300**: 48–50.
- Loy, A., Duller, S., Baranyi, C., Musmann, M., Ott, J., Sharon, I., *et al.* (2009) Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. *Environ Microbiol* **11**: 289–299.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., and Yadhukumar, B.A. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Martens, C.S. (1990) Generation of short chain acid anions in hydrothermally altered sediments of the Guaymas Basin, Gulf of California. *Appl Geochem* **5**: 71–76.
- Martin, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C., *et al.* (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* **24**: 1263–1269.
- Meyer, B., and Kuever, J. (2007) Phylogeny of the alpha and beta subunits of the dissimilatory adenosine-5'-phosphosulfate (APS) reductase from sulfate-reducing prokaryotes—origin and evolution of the dissimilatory sulfate-reduction pathway. *Microbiology* **153**: 2026–2044.
- Moran, M.A., Buchan, A., Gonzalez, J.M., Heidelberg, J.F., Whitman, W.B., Kiene, R.P., *et al.* (2004) Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**: 910–913.
- Moran, M.A., Satinsky, B., Gifford, S.M., Luo, H., Rivers, A., Chan, L.K., *et al.* (2013) Sizing up metatranscriptomics. *ISME J* **7**: 237–243.
- Musmann, M., Richter, M., Lombardot, T., Meyerdierks, A., Kuever, J., *et al.* (2005) Clustered genes related to sulfate respiration in uncultured prokaryotes support the theory of their concomitant horizontal transfer. *J Bacteriol* **187**: 7126–7137.
- Orcutt, B.N., Sylvan, J.B., Knab, N.J., and Edwards, K.J. (2011) Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol Mol Bio Rev* **75**: 361–422.
- Penn, K., Wu, D., Eisen, J., and Ward, N. (2006) Characterization of bacterial communities associated with deep-sea corals on Gulf of Alaska seamounts. *Appl Environ Microb* **72**: 1680–1683.
- Pereira, I.A., Ramos, A.R., Grein, F., Marques, M.C., da Silva, S.M., and Venceslau, S.S. (2011) A comparative

- genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. *Front Microbiol* **2**: 69, 1–22.
- Pham, V.D., Konstantinidis, K.T., Palden, T., and DeLong, E.F. (2008) Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environ Microbiol* **10**: 2313–2330.
- Pott, A.S., and Dahl, C. (1998) Sirohaem sulfite reductase and other proteins encoded by genes at the DSR locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur. *Microbiology* **144**: 1881–1894.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.
- Reinthal, T., and Van, H. (2010) Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep-Sea Res PT II* **57**: 1572–1580.
- Rich, V.I., Pham, V.D., Eppley, J., Shi, Y., and DeLong, E.F. (2011) Time-series analyses of Monterey Bay coastal microbial picoplankton using a 'genome proxy' microarray. *Environ Microbiol* **13**: 116–134.
- Sayavedra-Soto, L.A., Hamamura, N., Lui, C., Kimbrel, J.A., Chang, J.H., and Arp, D.J. (2011) The membrane-associated monooxygenase in the butane-oxidizing Gram-positive bacterium *Nocardioides* sp. strain CF8 is a novel member of the AMO/PMO family. *Environ Microbiol Rep* **3**: 390–396.
- Shi, Y., Tyson, G.W., and DeLong, E.F. (2009) Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459**: 266–269.
- Sorokin, D.Y. (2003) Oxidation of inorganic sulfur compounds by obligately organotrophic bacteria. *Microbiology* **72**: 641–653.
- Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Stamatakis, A., Hoover, P., and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* **57**: 758–771.
- Stockdreher, Y., Venceslau, S.S., Josten, M., Sahl, H., Pereira, I.A.C., and Dahl, C. (2012) Bacterium *Allochromatium vinosum*: evidence for sulfur transfer from DsrEFH to DsrC. *PLoS ONE*.
- Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., *et al.* (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the Dark Ocean. *Science* **333**: 1296–1300.
- Thauer, R.K., Jungerman, K., and Decker, K. (1977) Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol Rev* **41**: 100–180.
- Thorvaldsdóttir, H., Robinson, J., and Mesirov, J. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178–192.
- Toner, B.M., Fakra, S.C., Manganini, S.J., Santelli, C.M., Marcus, M.A., Moffett, J.W., *et al.* (2009) Preservation of iron(II) by carbon-rich matrices in a hydrothermal plume. *Nat Geosci* **2**: 197–201.
- Von Damm, K.L., Edmond, J.M., Measures, C.I., and Grant, B. (1985) Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California. *Geochim Cosmochim Acta* **49**: 2221–2237.
- Walsh, D., Zaikova, E., Howes, C., Song, Y., Wright, J., Tringe, S., *et al.* (2009) Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326**: 578–582.
- Welhan, J., and Lupton, J. (1987) Light hydrocarbon gases in Guaymas Basin hydrothermal fluids: thermogenic versus abiogenic origin. *Am Assoc Petr Geol B* **71**: 215–223.
- Winn, C.D., Karl, D.M., and Massoth, G.J. (1986) Microorganisms in deep-sea hydrothermal plumes. *Nature* **320**: 744–746.
- Wright, J., Konwar, K., and Hallam, S. (2012) Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* **10**: 381–475.
- Wright, T.D., Vergin, K.L., Boyd, P.W., and Giovannoni, S.J. (1997) A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl Environ Microb* **63**: 1441–1448.
- Zaikova, E., Walsh, D., Stilwell, C., Mohn, W., Tortell, P., and Hallam, S. (2010) Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* **12**: 172–191.

## Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Fig. S1.** Emergent self-organizing map based on tetranucleotide frequencies from metagenomic contigs greater than 2500 bp. Each data point represents a genomic fragment (2500–5000 bp), and the background topology represents the Euclidean distance between tetranucleotide frequency profiles. Thus genomic bins are delineated by ridges. The GB-SAR324 bin is identified, and other major genomic bins are coloured: Blue = MG1 Archaea, Dark Red = SUP05 and black = methylotrophs.

**Fig. S2.** Recruitment of genes from published single-cell SAR324 genomes to GB-Sar324 metagenome bin. Recruitment was performed using BLASTn with high stringency cut-offs (e-value cut-off 10<sup>-10</sup>, minimum length = 400 bases). Blue dots represent AFIB (AFIB00000000) and red dots represent AFIA (AFIA00000000) single-cell genomes. Blue and red lines represent the average nucleotide similarity to GB-SAR324.

**Fig. S3.** Recruitment of Guaymas Basin 454 metagenomic reads to the dark ocean SAR324 single amplified genomes. Reads were recruited with BLASTn using a similarity of > 90% and bit scores over 70. Coverage was calculated with a 100 base pair sliding window.

**Fig. S4.** Representative CTD depth profile from Guaymas Basin. A) Oxygen concentration, B) Beam transmission, C) Temperature and D) SAR324 abundance from clone libraries (Dick and Tebo, 2010) and high throughput tagged ribosomal sequencing libraries (Anantharaman, 2012). Note that the clone library data are as a percentage of bacteria, whereas the tag sequence data are as a percentage of bacteria plus archaea.

**Fig. S5.** Phylogenetic inference of large subunit 1,5-ribulose biphosphate genes (RuBisCO). SAR324-related sequences are highlighted in black and the GB-SAR324 RuBisCO gene is bolded.

**Fig. S6.** Recruitment of 454 metagenome reads to the *phmABC* operon from SAR324 single amplified genome (AAA240-J09). Reads were recruited with BLASTn using a similarity of > 90% and bit scores over 70.

**Fig. S7.** Recruitment of paired end Illumina metatranscriptome reads mapped to the *phmABC* operon from SAR324 single amplified genome (AAA240-J09). Reads were mapped with BWA and alignments were visualized with IGV (see *Methods* section). The order of genes for the *Phm* are CAB, which can clearly be seen through mapping.

**Fig. S8.** Depiction of contig synteny in GB-SAR324 and single-cell genome isolates. Contigs all contain sulfur cycling genes, the expression profiles of representative genes are shown in Fig. 3. These five contigs from GB-SAR324 are present in as single contig in the AFIA SAG (AAA240-J09), thus is used as a reference. Colours represent contigs from GB-SAR324, while inside contig boxes lines represent similarity between GB-SAR324 and SAGs.

**Fig. S9.** Phylogeny of the *dsrA* gene from SAR324 in relation to sulfur-oxidizing and sulfur-reducing micro-organisms.

**Fig. S10.** Phylogeny of the *aprA* gene from SAR324 to sulfur-oxidizing and sulfur-reducing micro-organisms.

**Fig. S11.** Phylogeny of the SAT gene from GB-SAR324 to sulfur-oxidizing and sulfur-reducing micro-organisms.

**Fig. S12.** Phylogeny of the SoxZ-like gene from SAR324 to sulfur-oxidizing micro-organisms.

**Fig. S13.** Phylogeny of the SoxY-like gene from GB-SAR324 to sulfur-oxidizing micro-organisms and phylogenetically similar desulfoferrodoxin genes.

**Fig. S14.** Phylogeny of the SoxB-like gene from GB-SAR324 to sulfur-oxidizing micro-organisms.

**Fig. S15.** Genes detected only in the background cDNA metatranscriptome. Percentages are calculated as fraction of genes found only in the background cDNA libraries.

**Fig. S16.** Genes detected only in the plume cDNA metatranscriptome. Percentages are calculated as fraction of genes found only in the plume cDNA libraries. The four unmarked categories at the bottom are related to (from left to right) quorum sensing and biofilm formation, macromolecule formation and synthesis, hormone synthesis and proteorhodopsin synthesis.

**Fig. S17.** Order of genes related to SAR324's putative *soxZ* and comparison to closest organisms in IMG determined by BLASTp. Genes that could potentially be related to the SOX system were translated by IMG and searched individually to find closest homolog proteins. Numbers between genes represent the percent similarity score from BLAST. Genes are colour coded for reference in closest relative genomes. The asterisk denotes that both the *soxZ* (30%) and the desulfoferrodoxin (35%) were both the top hits to the *Anaeromyxobacter* genome. SCGC AAA001-C10 is a dark ocean SAR324 single amplified genome. Lines behind genes denotes whether genes were found on a single (no breaks) or multiple contigs and the number at the beginning and end denote region of the genome.

**Table S1.** Assessment of genome completeness.

**Table S2.** Abundance of core metabolimic pathway genes in GB-SAR324.

**Table S3.** Annotation and abundance of Genomic Bin SAR324.