Title: Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria

Authors: Karthik Anantharaman[1], John A. Breier[2], Cody S. Sheik[1] & Gregory J. Dick[1,3]

Affiliations:

1. Dept. of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI 48109, USA.
2. Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.
3. Dept. of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA.

Correspondence and requests for materials should be addressed to Gregory J. Dick (Email: gdick@umich.edu)

CLASSIFICATION : BIOLOGICAL SCIENCES; MICROBIOLOGY

Manuscript information: Submitted to PNAS as a Research Report
Number of pages (including reference and figure legends): 15
Number of tables: 0
Number of figures: 4
Supporting Information: 29 pages, 6 supplementary tables + 8 supplementary figures

Keywords: Metagenomics, metatranscriptomics, SUP05, hydrogen oxidation, deep oceans

\abstract

Hydrothermal vents are a well-known source of energy that powers chemosynthesis in the deep sea. Recent work suggests that microbial chemosynthesis is also surprisingly pervasive throughout the dark oceans, serving as a significant CO_2 sink even at sites far-removed from vents. Ammonia and sulfur have been identified as potential electron donors for this chemosynthesis, but they do not fully account for measured rates of dark primary production in the pelagic water column. Here we use metagenomic and metatranscriptomic analyses to show that deep-sea populations of the SUP05 group of uncultured sulfur-oxidizing *Gammaproteobacteria*, which are abundant in widespread and diverse marine environments, contain and highly express genes encoding group 1 Ni-Fe hydrogenase enzymes for H_2 oxidation. Reconstruction of near-complete genomes of two co-occurring SUP05 populations in hydrothermal plumes and deep waters of the Gulf of California enabled detailed population-specific metatranscriptomic analyses, revealing dynamic patterns of gene content and transcript abundance. SUP05 transcripts for genes involved in H_2 and sulfur oxidation are most abundant in hydrothermal plumes where these electron donors are enriched. In contrast, a second hydrogenase has more abundant transcripts in background deep sea samples. Coupled with results from a bioenergetic model that suggest that H_2 oxidation can contribute significantly to the SUP05 energy budget, these findings reveal the potential importance of H_2 as a key energy source in the deep ocean. This study also highlights the genomic plasticity of SUP05, which enables this widely distributed group to optimize its energy metabolism (electron donor and acceptor) to local geochemical conditions.

\body

Deep-sea hydrothermal vent ecosystems depend on microorganisms that utilize reduced chemicals such as sulfur, methane, ammonium, and hydrogen (H_2) as electron donors for chemosynthesis (*1-5*). Recent work suggests that microbial chemosynthesis is also far more prevalent in the broader deep oceans than previously recognized, extending throughout the water column of the dark open ocean, where it serves as a significant source of organic carbon *(6, 7*). The fuels for this pelagic primary production remain unknown, but recent studies show that ammonium (*3*) and sulfur *(8, 9*) are potential electron donors in the water column. Hydrogen (H_2), long known as an energy source for free-living bacteria in seafloor hydrothermal systems, was also recently identified as an electron donor in hydrothermal vent animal symbioses (*4*). Although microbial communities at seafloor hydrothermal vent sites have attracted much attention, hydrothermal vent plumes remain poorly characterized despite their importance as habitats for free-living chemolithoautotrophs *(10).* These plume microorganisms mediate the hydrothermal transfer of elements from the lithosphere to the oceans *(11, 12)* and contribute significantly to organic carbon in the deep oceans via carbon fixation *(1, 13-15).*

We investigated hydrothermal vent plumes in Guaymas Basin (GB) where hydrothermal enhancement of microbial activity is evident through increased total RNA concentrations *(16)* and rapid microbially-catalyzed Mn oxidation rates *(11)* in comparison to background waters of the deep Gulf of California. Among the most active and abundant microorganisms in GB plumes are sulfur-oxidizing bacteria of the SUP05 group of *Gammaproteobacteria (13, 16).* SUP05 are dominant members of microbial communities in diverse marine environments such as hydrothermal vent plumes, symbiotic associations with hydrothermal vent clams and mussels, and oxygen minimum zones (OMZ) across the world`s oceans *(9, 17-22).*

In the present study, we use a combination of DNA, cDNA, SSU rRNA amplicon sequencing, and thermodynamic/bioenergetic modeling to elucidate the genetic potential, transcriptional activity and distribution of two uncultivated lineages of SUP05 bacteria in hydrothermal plumes and surrounding deep-sea waters. We report evidence for H_2 oxidation as an important source of electrons for microbial growth in the deep oceanic water column and suggest that the SUP05 group displays metabolic plasticity that underlies the phylogenetic diversity of these widespread bacteria.

Results and Discussion

Diversity and distribution of SUP05 at Guaymas Basin. Phylogenetic analysis of SUP05 small subunit (SSU) rRNA gene sequences from Guaymas Basin hydrothermal plumes revealed the presence of two distinct SUP05 populations (Fig. S1) (hereby referred to as GB-1 and GB-2) that share 96.7% SSU rRNA nucleotide sequence identity *(16)*. Our analyses also show that these two SUP05 lineages cluster closely with all previously identified SUP05 populations and fall into two co-occurring distinct sub-clades, similar to sequences retrieved from the African Shelf Namibian Upwelling zone and from the Saanich Inlet oxygen minimum zone (OMZ). The closest relatives of the GB SUP05 are the SUP05 SI-1 lineage (GB-1) *(9)* and symbionts of *Bathymodiolus* mussels from hydrothermal vents (GB-2) *(23, 24)*. High-throughput sequencing of the SSU rRNA gene amplicons from the Guaymas Basin water column indicate that GB-1 & 2 dominate the deep waters of the GB (>1700m), comprising up to 30% of the microbial community (Fig. S2). The abundance of SUP05 is tightly coupled to hydrothermal signals and also shows a minor increase in the oxygen minimum zone of the upper GB water column (Fig. S2).

Recovery and comparative analysis of SUP05 genomes. High-throughput sequencing of community genomic DNA and cDNA was used to reconstruct the metagenomes and metatranscriptomes of GB-1 and GB-2 in hydrothermal plumes and surrounding waters of the deep Gulf of California. *De novo* metagenomic assembly and binning by tetranucleotide signatures (Fig. S3) and BLAST (Supplementary Methods) yielded draft genomes of GB-1 & 2 that span 1.24 and 1.26 million base pairs (Mbp) of consensus sequence respectively, with an average coverage of ~13x for both genomes (Table S2). To confirm that they represented near-complete genomes, we identified a complete set of universally conserved genes present in each SUP05 genome (Table S3).

GB-1 & 2 shared 83% of predicted genes with each other, and 60% of predicted genes with SUP05 populations from the Saanich Inlet OMZ (*9*) and the clam symbionts, *Candidatus* Ruthia Magnifica *(18)* and *Candidatus* Vesicomyosocius okutanii *(21)* (Fig. 1). Like other SUP05 populations sequenced to date, GB-1 & 2 possess the complete repertoire of genes for carbon fixation and oxidation of reduced sulfur compounds, consistent with a common sulfur-based chemolithoautotrophic metabolism within the SUP05 group. These genes encode enzymes for

the oxidation of reduced sulfur compounds (H_2S, $S_2O_3^{2-}$, S^0, SO_3^{2-}) including sulfide quinone oxidoreductase (*sqr*), mediating the oxidation of sulfide (HS^-) to elemental sulfur (S^o), the Sox enzyme complex (*soxABXYZ*) for oxidation of thiosulfate ($S_2O_3^{2-}$) to elemental sulfur, rhodanese sulfurtranferase for oxidation of thiosulfate to sulfite, reverse dissimilatory sulfite reductase complex (*dsrAB*) for oxidation of elemental sufur to sulfite (SO_3^{2-}), adenosine 5'-phosphosulfate reductase (*aps*) and sulfate adenylyltransferase (*sat*) for oxidation of sulfite to sulfate (SO_4^{2-}) (Fig. 2). Absence of *soxCD* genes in SUP05 populations may result in storage of elemental sulfur and provisioning of SUP05 with an electron donor *(25)*, similar to the recently cultivated heterotrophic ARCTIC96BD-19 clade bacterium *(26)*.

Metabolic plasticity in SUP05 – genes for H_2 oxidation and O_2 respiration. The Guaymas Basin SUP05 populations also harbor genes that set them apart from their Saanich Inlet and clam symbiont counterparts (Fig. 1). Key among these unique genes is a membrane-bound group 1 Ni-Fe hydrogenase for H_2 oxidation *(27)*. This enzyme and its associated maturation factors are encoded in both SUP05 populations by a set of 18 genes, 17 of which are adjacent on contigs (Fig. 3) confidently assigned to SUP05 by tetranucleotide frequency and by the fact that genes flanking the hydrogenase operon share synteny and high sequence similarity with other SUP05 genomes. Although this GB-SUP05 hydrogenase is not present in the Saanich Inlet OMZ SUP05 *(9)* or the clam symbionts *Candidatus* Ruthia Magnifica *(18)* and *Candidatus* Vesicomyosocius okutanii *(21)*, it is phylogenetically affiliated with other hydrothermal vent-derived hydrogenases (Fig. 4), including those from recently discovered H_2-oxidizing symbionts of *Bathymodiolus* mussels that are the first known H_2-powered chemosynthetic symbiosis at deep-sea hydrothermal vents (*4*). Genes in the SUP05 hydrogenase operons display synteny and high sequence identity (92 and 94% for HupS & HupL) with genes from the *Bathymodiolus* symbionts for structural assembly, synthesis, hydrogen uptake and oxidation, suggesting a similar role in H_2 oxidation for the purpose of energy production (Fig. 3A).

The SUP05 genomic bin also contains a contig (AJXC01001965) with genes encoding a second group I Ni-Fe hydrogenase that displays distinct operon structure (Fig. 3B) and phylogeny (Fig. 4) to the first. This putative SUP05 hydrogenase clusters with Ni-Fe hydrogenases from epipelagic *Gammaproteobacteria (28)*, *Flavobacteria (29, 30)*, and *Deltaproteobacteria (31)*, possibly indicating a different evolutionary origin and/or physiological role. Because the second

148 hydrogenase-containing contig cannot be scaffolded onto other SUP05 genomes, and in view of its complex phylogeny, we cannot conclusively determine the taxonomic origin at this time.

150 The Guaymas Basin SUP05 genomes also display genomic and metabolic diversity in terms of electron acceptors for energy metabolism. GB-1 & 2 genomes encode for both a cytochrome c 152 oxidase and a cbb_3-type terminal cytochrome c oxidase. Both these cytochrome c oxidase complexes are shared by the SUP05 clam symbionts *(18, 21)* but are absent in the free-living 154 OMZ SUP05 (*9*). The presence of these genes enables the use of oxygen as a terminal electron acceptor in both oxic and microoxic environments that exist in the stratified water column of 156 Guaymas Basin. The set of genes encoding for dissimilatory nitrate/nitrite reduction to N_2O, which are present in OMZ SUP05, are absent in GB-1 & 2 (Supporting Information) except for a 158 single dissimilatory nitrite reductase (NO-forming *nirK*) in GB-2, hinting at a possible role in either a partial dissimilatory denitrification pathway or in nitrite detoxification rather than the 160 full denitrification pathway of Saanich Inlet SUP05 *(9)*. Evidence consistent with such partial denitrification has been found in the Eastern Tropical South Pacific OMZ, where SUP05 are 162 abundant, and sulfide-dependent reduction of nitrate produces NO_2^- and N_2O as well as N_2 *(17)*. Genes for dissimilatory nitrate and nitrite reductases and associated cofactors were identified on 164 short contigs with low genomic coverage (Table S4) suggesting that they stem from minor genome variants within the community. Genes for reduction of nitric oxide (NO) to nitrous oxide 166 (N_2O) (*norB*, *norC*) were absent from the metagenome.

Carbon and nitrogen metabolism. The GB-1 & 2 genomes contain genes encoding the Calvin-168 Benson-Bassham (CBB) cycle including a single form II ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) for the purpose of carbon fixation. This form II RuBisCO is also present 170 in the Saanich Inlet OMZ SUP05 *(9)* and the clam symbionts, *Candidatus* Ruthia Magnifica *(18)* and *Candidatus* Vesicomyosocius okutanii *(21)*. In contrast, the H_2-oxidizing symbionts of 172 *Bathymodiolus* mussels *(4)* possess genes for a form I RuBisCO, which is optimized for higher O_2 and lower CO_2 concentration *(32)*. The presence of genes in GB SUP05 encoding for form II 174 RuBisCO enzymes typically adapted to low O_2 and high CO_2 concentrations is consistent with the low O_2 conditions of the deep Guaymas Basin. Genes for gluconeogenesis and the non-176 oxidative branch of the pentose phosphate pathway were also identified, along with all components of the tricarboxylic acid cycle (TCA) except for those encoding the α-ketoglutarate

dehydrogenase enzyme, consistent with GB SUP05 being primarily autotrophs *(9)*. This is also evidenced by the lack of known transporters for organic carbon except the two noted below.

The GB SUP05 genomes possess two ABC-type transporters (HAAT and PAAT family) annotated as amino acid transporters. GB-2 also contains a single putative di/tri carboxylate transporter. The presence of these transporters is intriguing because it may suggest an alternative source of carbon and nitrogen and hint at a mixotrophic lifestyle with the ability to utilize organic carbon as in the recently cultivated and closely related ARCTIC96BD-19 clade bacterium *(8, 26)*. For the purpose of nitrogen assimilation and metabolism, GB-1 & 2 genomes have multiple copies of genes for ammonium transport and a full complement of assimilatory nitrite reduction genes for reduction of nitrite (NO_3^-) to ammonia (NH_3). Also present are genes for breakdown of urea and amides by an amidohydrolase (GB-1 & 2) and a urease (GB-1 only) that are absent in the SUP05 clam symbionts and the Saanich Inlet OMZ SUP05 genomes.

Population-specific metatranscriptomic mapping. In order to examine the transcriptional activity of GB-1 & 2, we used their assembled genomes as a framework to map metatranscriptomic reads. Population-specific mapping of Illumina cDNA reads to both SUP05 genomes assigned a total of 104,075 transcripts to GB-1 and 136,524 transcripts to GB-2. Both of these SUP05 genomes recruited more total transcripts in the hydrothermal plume than background by a ratio of approximately three, indicating that they are stimulated in hydrothermal plumes. Amongst the most abundant transcripts in the metatranscriptome were those mapping to genes involved in chemolithoautotrophy (Fig. S4), including H_2 oxidation (Fig 3C, 3D), O_2 respiration (Fig. S5), oxidation of reduced sulfur species (Fig. 2), and carbon fixation (Fig. S6). Both GB-1 & 2 preferentially expressed genes for oxidation of multiple reduced sulfur species (H_2S, $S_2O_3^{2-}$, S^0, SO_3^{2-}) in the plume compared to the background, indicating that reduced sulfur species are important electron donors in the plume. High transcript abundances of the RuBisCO genes in both GB-1 & 2 metatranscriptomes implicates the deep-sea SUP05 populations in carbon fixation and underscores their importance as key autotrophs in the deep sea. All SUP05 genes for nitrogen metabolism were recovered in the metatranscriptome, with genes encoding ammonium and amino acid/amide uptake having high transcript abundances in both the hydrothermal plume and background deep-sea, again suggesting that GB SUP05 actively obtain

amino acids from the environment (Fig. S7). These trends in transcript abundance for SUP05 genes were similar across both 454 and Illumina-based metatranscriptomes.

Dynamic expression of Ni-Fe hydrogenase genes. A major difference is evident in patterns of hydrogenase transcript abundance between plume and background. The hydrothermal vent-related hydrogenases are highly enriched in metatranscriptomes of plumes (Fig. 3C). Conversely, the epipelagic-related hydrogenase (putative SUP05) is enriched in the background metatranscriptome relative to the plume (Fig. 3D). Because SUP05 abundance in the metagenome is similar between plume and background *(13, 16)*, the dynamic patterns of transcript abundance we observe for the Ni-Fe hydrogenase genes suggests that their expression is regulated rather than constitutive *(4)*. Based on the increased H_2 concentrations expected in plumes versus background, we suggest that H_2 concentration is the likely regulator of this observed differential expression. We speculate that the hydrothermal vent-related and the epipelagic-related hydrogenases are distinguished in their affinity for H_2, the former being adapted to higher H_2 concentrations in environments such as hydrothermal plumes, and the latter to low H_2 concentrations typically available in the background deep ocean.

As abundant members of both hydrothermal plume and background deep ocean communities, SUP05 populations likely take advantage of H_2 derived not only from hydrothermal fluids but also from mineral precipitation reactions in the plume *(15)* and possibly anaerobic decomposition on sinking particles, a source of H_2 posited long ago *(33)*. Further, high levels of expression of Ni-Fe hydrogenases in the background deep ocean, far from the hydrothermal plumes, may also indicate the presence of a significant but currently unrecognized source of H_2.

Thermodynamic model for estimation of plume chemistry and bioenergetics. H_2 concentrations of up to 3 mM measured in GB end-member hydrothermal fluids are the result of the reaction of seawater with mantle-derived basalt in the oceanic crust at high temperature and pressure *(34-36)*. Unfortunately, to our knowledge, no direct measurements of H_2 concentration have been made on GB plumes. Thus, we used equilibrium thermodynamic reaction path modeling to estimate the concentrations of H_2 and other potential electron donors in the GB plume (Supplementary Information). Results predict that H_2 concentrations range from 0.5 to 50 nM in plumes sampled here (2.93-2.97^{o}C), which are up to ~100 times greater than typical H_2 concentrations of 0.4 nM in the background deep sea.

To assess the relative importance of H_2 and sulfur as energy sources for SUP05, we compared the free energy yields for a number of metabolisms including those using H_2, S^0, H_2S, $S_2O_3^{2-}$ and particulate metal sulfides as electron donors (Table S6). Our model estimated the free energy available from H_2 oxidation in the hydrothermal plume to be 0.04 J/kg of plume fluid at a temperature of 2.95°C, representing 17% of the energy budget for SUP05 (Fig. S8). Further, these results indicate that H_2 oxidation can account for up to 22% of the energy budget of SUP05 in warmer fluids of rising hydrothermal plumes (3.0-5.9°C), which have not yet been studied from a microbiological perspective. This prominent role for H_2 oxidation is consistent with previous studies that have modeled available energy in hydrothermal plumes *(15)*; H_2 oxidation is expected to play an even more important role in ultramafic-hosted hydrothermal systems *(37)*. Amongst sulfur species, we found S^0 oxidation with both oxygen and nitrate to be thermodynamically favored relative to H_2S, thiosulfate and particulate metal sulfides. Although there is uncertainty with regard to sulfur speciation in the plume and the actual form of sulfur utilized by SUP05 is unknown, these results suggest that Guaymas Basin SUP05 populations utilize environmentally-supplied sulfur species other than dissolved H_2S. Overall, the modeling results presented here indicate that oxidation of H_2 and reduced sulfur species are both potentially significant sources of free energy for growth of SUP05 populations in Guaymas Basin hydrothermal plumes.

Conclusions.

This study advances our understanding of the chemolithotrophic metabolism of a widespread group of marine bacteria, providing insight into potential genetic and physiological underpinnings and biogeochemical implications of microbial diversity observed within the SUP05 group. As abundant microorganisms in the pelagic realm of the dark ocean, SUP05 have the capacity to influence and link the global cycles of sulfur, nitrogen, and carbon in an environment that holds the largest reservoir of reactive dissolved inorganic carbon on the Earth's surface. Recognition of H_2 as a significant electron donor for microbial growth in the pelagic water column may shed light on discrepancies in current oceanic carbon budgets *(6, 7)*. Additional molecular studies are needed to determine the prevalence of SUP05 hydrogen oxidation genes beyond the Gulf of California (Supporting Information), and geochemical measurements of H_2 oxidation rates are required to directly and quantitatively evaluate the

contribution of H_2 to chemosynthesis in the deep sea. Although these experiments are challenging due to the low H_2 concentrations (nM) and remote nature of the deep sea, the molecular evidence presented here provides the impetus to develop such methods. The genetic and metabolic plasticity of electron donors (H_2 and reduced sulfur species) and acceptors (oxygen, nitrate, and nitrite) across the SUP05 group revealed here underscores the importance of taking fine-scale microbial functional diversity into account when tracking microbial biogeochemistry. Given the central role of SUP05 in the biogeochemistry of globally expanding OMZs and associated feedbacks on cycling of carbon, nitrogen, sulfur, and greenhouse gases (*20*), such resolution will be critical to understanding and predicting marine ecosystem dynamics in the context of environmental change.

Methods

Sampling. Samples were collected on three cruises aboard *R/V New Horizon* in 2004 and 2005 as described previously *(11, 13)*. Metadata and chemical/physical characteristics of samples used for shotgun DNA and cDNA sequencing are presented in detail in Lesniewski et al. 2012 *(16)*, while summaries of these samples along with those used for SSU rRNA gene amplicon sequencing are described in Table S1.

Extraction of nucleic acids, metagenomic and metatranscriptomic sequencing. DNA and RNA extraction were done as described previously *(13, 16)*. Purified DNA was used to prepare DNA libraries for sequencing using standard protocols (454 Life Sciences). An overall summary of DNA sequencing obtained using 454 GS FLX Titanium is presented in Lesniewski et al. 2012 *(16)*. cDNA synthesis was performed as described previously *(38)*. cDNA sequencing produced a total of 1,558,905 reads from the plume (664,240 from Plume-3 (Cast 21-6#2) and 894,665 from Plume-4 (Cast 12-27a#1)) and 1,008,693 reads from the background deep sea (514,607 from Background-1 (Cast 12-8#12) and 504,086 from Background-2 (Cast 34-2#7)) using 454 GS FLX Titanium. A plume and background cDNA sample each were prepared for resequencing (for the purpose of comparison with 454) using standard protocols (Illumina) and a total of 103,078,758 reads from the plume (Cast 21-6#2, Plume-3) and 122,259,588 reads from the background deep sea (Cast 12-8#12, Background-1) were obtained using Illumina HiSeq2000.

Assembly and Annotation. *De novo* metagenomic assembly was performed as described previously *(16)* using MIRA *(39)* with parameters as follows: (-job=denovo, genome, accurate, 454 -notraceinfo -CL:pec=no -GE:not=8 -AS:urd=no -SK:bph=12:pr=80 454_SETTINGS -AS:mrl=50 -CO:mrpg=3 -AL:mrs=80). Gene annotations of assembled contigs was done through Integrated Microbial Genomes & Metagenomics (IMG/M) system *(40)* as described previously *(16)*. See *SI Appendix* for information on binning, identification and separation of the SUP05 contigs.

cDNA mapping. Transcript reads were mapped to predicted proteins using BLASTN *(bitscore≥50, E≤ 1 x E^{-5}, percent identity≥95%)*. Numbers of hits per gene were normalized by dividing the total cDNA hits by gene length, multiplying by 1000 and adjusting for the total size of the data set to enable comparison across the multiple data sets in the background deep sea and hydrothermal plume. Trends in normalized transcript abundances were similar across both 454 and Illumina data sets.

SSU rRNA gene amplicon pyrosequencing. DNA was extracted from a ¼ filter with the MoBio PowerSoil DNA isolation kit (Carlsbad, CA, USA). In addition to bead beating, filters were incubated at 65^{o}C for 20 min to facilitate cellular lysis. Bead beating was performed using the MP-Bio FastPrep-24 (Santa Ana, CA, USA) for 45 seconds at setting 6.5. The 16S rRNA gene was amplified in triplicate 25 μL reactions containing the following (final concentration): 12.5 μL 5 Prime HotMasterMix (Gaithersburg, MD, USA), 2 μL (15 μM) each forward and reverse primers, 1 μL community DNA. Previously described 16S rRNA gene primers targeting the V4 region (515F/806R) *(41)* were used and the reverse primers contained a 12-base barcode *(42)*. PCR thermocycler conditions were as follows: initial denaturation 95^{o}C -4 min followed by 30 rounds of 95^{o}C for 30 sec, 50^{o}C for 1 min, 72^{o}C for 1 min and final elongation 72^{o}C for 10 min. Triplicate PCRs were combined and cleaned using a MoBio UltraClean PCR Clean-up kit (Carlsbad, CA, USA). DNA concentration was quantified using PicoGreen (Invitrogen, Carlsbad, CA, USA). Individual barcoded samples were combined into a single sample at equivalent concentrations then sent to Engencore (http://engencore.sc.edu) for pyrosequencing using 454 Titanium chemistry. Amplicon reads were corrected with Pyronoise *(43)* implemented in MOTHUR (v. 1.26.0)*(44)*. Operational taxonomic units (OTUs) were binned at 99% similarity and chimera checked using the OTUpipe (http://drive5.com/otupipe) command within Qiime

(ver 1.4.0) (45). Default parameters were used with the exceptions of initial clustering at 100% similarity and low abundance OTUs being kept for downstream analysis of rare phylotypes. OTUs were taxonomically classified with BLASTn (ver 2.2.22, e-values cutoff 10^{-8}) using Greengenes taxonomy and fasta files (available at http://qiime.wordpress.com), which were customized to include SUP05 16S rRNA sequences recovered from Guaymas Basin metagenomic libraries. Binning of OTUs at 99% was necessary in order to distinguish the two SUP05 phylotypes (GB-1 and GB-2). Using the full-length 16S rRNA genes recovered from the metagenomic libraries, we determined that for the V4 region used in the pyrosequencing study, an OTU cutoff of >98.5% would be necessary to distinguish the GB-1 & 2 phylotypes.

Acknowledgements. This project is funded in part by the Gordon and Betty Moore Foundation and the National Science Foundation (OCE 1029242). We also thank the University of Michigan Rackham Graduate School Faculty Research Fellowship Program for their support and Brett Baker and Sunit Jain for their insightful assistance. 454 DNA sequencing was conducted by Lynn Tomsho in the laboratory of Stephan Schuster at Penn State University. Illumina sequencing was conducted at the University of Michigan DNA Sequencing Core.

Accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AJXC00000000. The version described in this paper is the first version, AJXC01000000.

References

1. de Angelis MA, Lilley MD, & Baross JA (1993) Methane oxidation in deep-sea hydrothermal plumes of the endeavour segment of the Juan de Fuca Ridge. *Deep Sea Research Part I: Oceanographic Research Papers* 40(6):1169-1186.
2. Distel DL, *et al.* (1988) Sulfur-oxidizing bacterial endosymbionts: analysis of phylogeny and specificity by 16S rRNA sequences. *J Bacteriol* 170(6):2506-2510.
3. Lam P, Cowen JP, & Jones RD (2004) Autotrophic ammonia oxidation in a deep-sea hydrothermal plume. *FEMS Microbiol Ecol* 47(2):191-206.
4. Petersen JM, *et al.* (2011) Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* 476(7359):176-180.
5. Jannasch HW & Mottl MJ (1985) Geomicrobiology of Deep-Sea Hydrothermal Vents. *Science* 229(4715):717-725.
6. Aristegui J, Gasol JM, Duarte CM, & Herndl GJ (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnol. Oceanogr.* 54(5):1501-1529 .
7. Reinthaler T, van Aken HM, & Herndl GJ (2010) Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep Sea Research Part II: Topical Studies in Oceanography* 57(16):1572-1580.

8. Swan BK, *et al.* (2011) Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* 333(6047):1296-1300.
9. Walsh DA, *et al.* (2009) Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 326(5952):578-582 .
10. Winn CD, Karl DM, & Massoth GJ (1986) Microorganisms in deep-sea hydrothermal plumes. *Nature* 320(6064):744-746.
11. Dick GJ, *et al.* (2009) Enzymatic microbial Mn(II) oxidation and Mn biooxide production in the Guaymas Basin deep-sea hydrothermal plume. *Geochimica et Cosmochimica Acta* 73(21):6517-6530.
12. Toner BM, *et al.* (2009) Preservation of iron(II) by carbon-rich matrices in a hydrothermal plume. *Nature Geosci* 2(3):197-201.
13. Dick GJ & Tebo BM (2010) Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ Microbiol* 12(5):1334-1347 .
14. Lilley MD, Feely, R.A., and Trefry, J.H. ed (1995) *Chemical and biochemical transformations in hydrothermal plumes* (American Geophysical Union, Washington, DC, USA), pp 369–391.
15. McCollom T (2000) Geochemical constraints on primary productivity in submarine hydrothermal vent plumes. *Deep Sea Research Part I: Oceanographic Research Papers* 47(1):85-101.
16. 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*.
17. Canfield DE, *et al.* (2010) A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* 330(6009):1375-1378 .
18. Newton IL, *et al.* (2007) The Calyptogena magnifica chemoautotrophic symbiont genome. *Science* 315(5814):998-1000 .
19. 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(2):1190-1198 .
20. Wright JJ, Konwar KM, & Hallam SJ (2012) Microbial ecology of expanding oxygen minimum zones. *Nat Rev Micro* 10(6):381-394.
21. Kuwahara H, *et al.* (2007) Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, Calyptogena okutanii. *Curr Biol* 17(10):881-886 .
22. Lavik G, *et al.* (2009) Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* 457(7229):581-584.
23. Duperron S, *et al.* (2006) A dual symbiosis shared by two mussel species, Bathymodiolus azoricus and Bathymodiolus puteoserpentis (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environ Microbiol* 8(8):1441-1447.
24. Petersen JM, Wentrup C, Verna C, Knittel K, & Dubilier N (2012) Origins and Evolutionary Flexibility of Chemosynthetic Symbionts From Deep-Sea Animals. *Biol Bull* 223(1):123-137.
25. Hensen D, Sperling D, Trüper HG, Brune DC, & Dahl C (2006) Thiosulphate oxidation in the phototrophic sulphur bacterium Allochromatium vinosum. *Mol Microbiol* 62(3):794-810.
26. Marshall KT & Morris RM (2012) Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J*.
27. Vignais PM & Billoud B (2007) Occurrence, Classification, and Biological Function of Hydrogenases: *Ann Rev Microbiol* 107(10):4206-4272.
28. Huggett MJ & Rappé MS (2012) Genome Sequence of Strain HIMB30, a Novel Member of the Marine Gammaproteobacteria. *J Bacteriol* 194(3):732-733.
29. Woyke T, *et al.* (2009) Assembling the Marine Metagenome, One Cell at a Time. *PLoS One* 4(4):e5299.
30. Cho J-C & Giovannoni SJ (2004) Robiginitalea biformata gen. nov., sp. nov., a novel marine bacterium in the family Flavobacteriaceae with a higher G+C content. *Int J Syst Evol Microbiol* 54(4):1101-1106.

31. Chitsaz H, *et al.* (2011) Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nat Biotech* 29(10):915-921.
32. Badger MR & Bek EJ (2008) Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *J Exp Bot* 59(7):1525-1541.
33. Karl DM, Knauer GA, Martin JH, & Ward BB (1984) Bacterial chemolithotrophy in the ocean is associated with sinking particles. *Nature* 309(5963):54-56.
34. Lilley MD, de Angelis MA, & Gordon LI (1982) CH4, H2, CO and N2O in submarine hydrothermal vent waters. *Nature* 300(5887):48-50.
35. McCollom TM ed (2008) *Observational, experimental, and theoretical constraints on carbon cycling in mid-ocean ridge hydrothermal systems* (AGU, Washington, D. C.), Vol 178, pp 193-213.
36. Welhan JA & Craig H (1979) Methane and Hydrogen in East Pacific Rise Hydrothermal Fluids. *Geophysical Research Letters* 6(11):829-831 .
37. Amend JP, McCollom TM, Hentscher M, & Bach W (2011) Catabolic and anabolic energy for chemolithoautotrophs in deep-sea hydrothermal systems hosted in different rock types. *Geochimica et Cosmochimica Acta* 75(19):5736-5748.
38. Frias-Lopez J, *et al.* (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci* 105(10):3805-3810.
39. Chevreux B (2005) MIRA: An Automated Genome and EST Assembler Ph.D (German Cancer Research Center Heidelberg, Duisburg).
40. Markowitz VM, *et al.* (2008) IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res* 36(suppl 1):D534-D538.
41. Bates ST, *et al.* (2010) Examining the global distribution of dominant archaeal populations in soil. *ISME J* 5(5): p. 908-917.
42. Fierer N, Hamady M, Lauber CL, & Knight R (2008) The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci* 105(46):17994-17999.
43. Quince C, *et al.* (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Meth* 6(9):639-641.
44. Schloss PD, *et al.* (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* 75(23):7537-7541.
45. Caporaso JG, *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* 7(5):335-336.

Foot notes

Correspondence and request for materials should be addressed to G.J.D. (gdick@umich.edu).

Author Contributions: G.J.D. collected the samples. K.A. and G.J.D. designed the study. G.J.D. and K.A did the DNA and cDNA sequencing. K.A. did the data analyses. J.A.B. did the thermodynamic modelling. C.S.S. did the SSU rRNA amplicon sequencing. K.A. and G.J.D. wrote the manuscript. K.A., J.A.B and C.S.S wrote the supplementary information.

The authors declare no conflict of interest.

This article contains supporting information online at www.pnas.org/

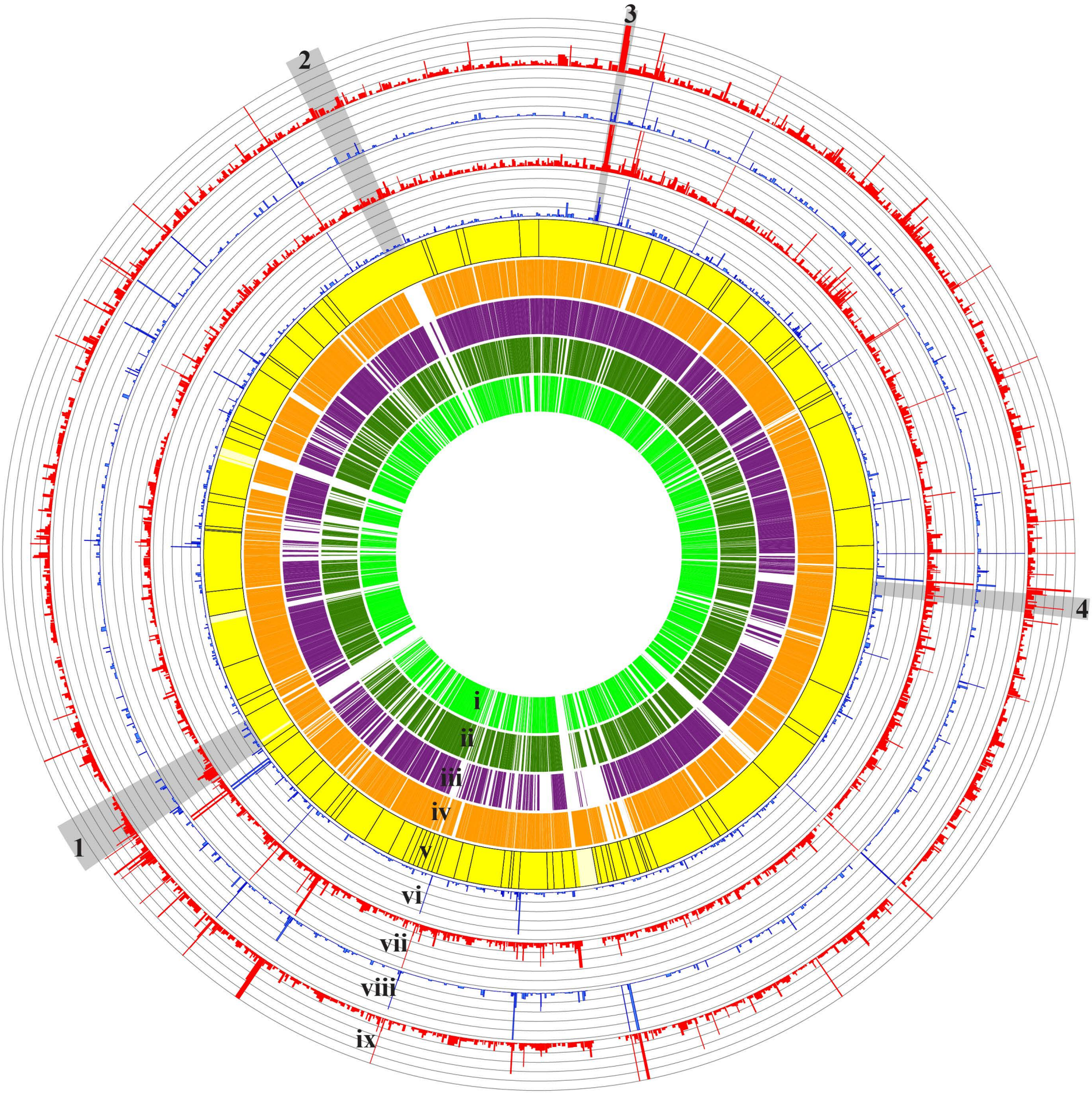
FIGURE LEGENDS

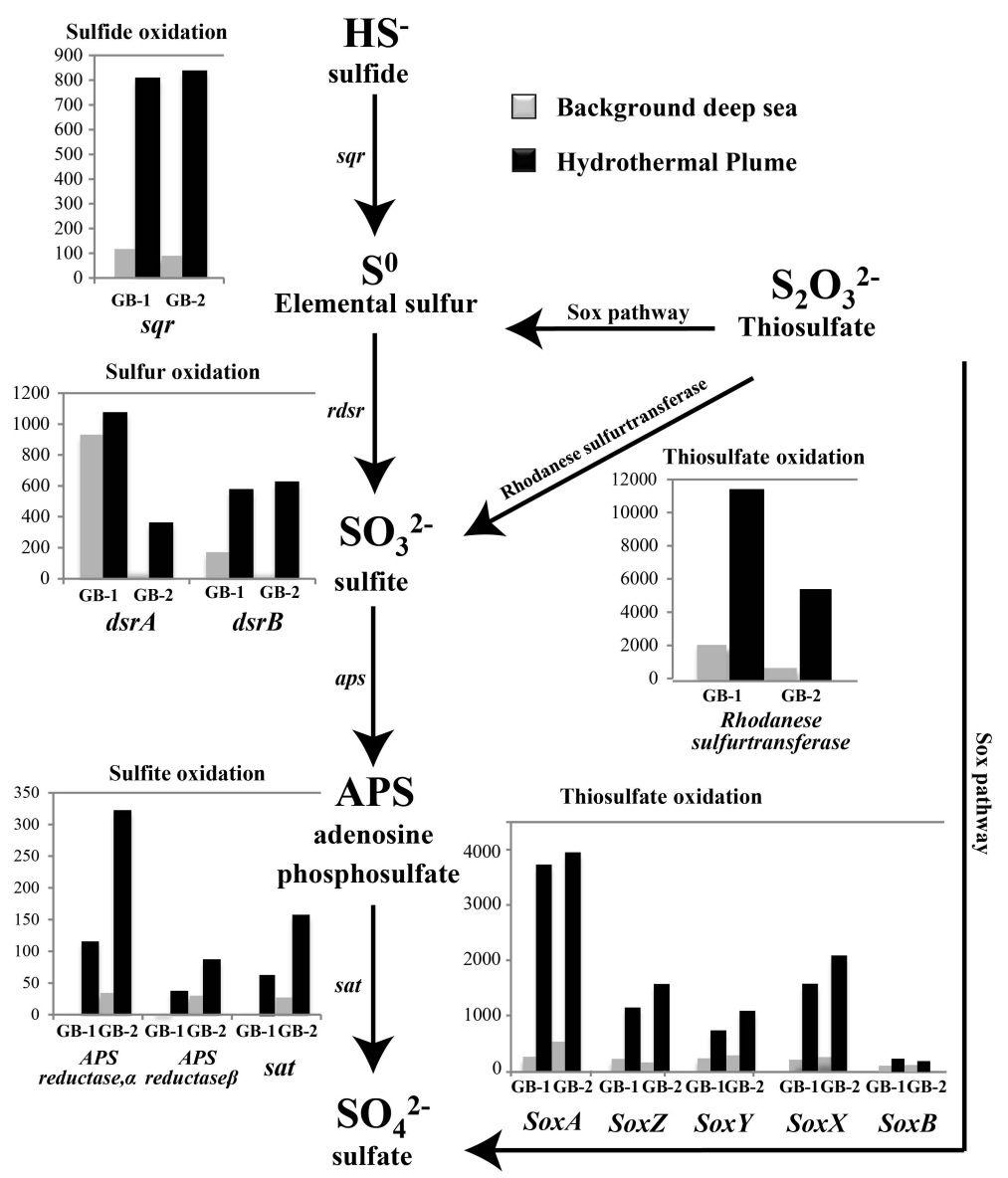
Figure 1. Content and transcript abundance of genes from Guaymas Basin SUP05 populations and comparison to genomes of other sequenced SUP05. Nested circles from innermost to outermost represent: **(i) – (v)** gene content with reference to GB-1 – (i) *Candidatus* Vesicomyosocius okutanii; (ii) *Candidatus* Ruthia magnifica; (iii) Saanich Inlet OMZ SUP05; (iv) GB-2; (v) GB-1. Gaps indicate the absence of genes in comparison to other SUP05 genomes. Black lines on GB-1 denote the separation of contigs that comprise the metagenome. **(vi) – (ix)** normalized abundance of 454 transcripts: (vi) GB-2 transcripts in background (blue); (vii) GB-2 transcripts in plume (red); (viii) GB-1 transcripts in background (blue); (ix) GB-1 transcripts in plume (red). Grey highlights on outermost circles indicate genes of interest: 1 - hydrogenase operon; 2 – urease operon; 3 – *sox* operon; 4 – cytochrome c oxidase complex.

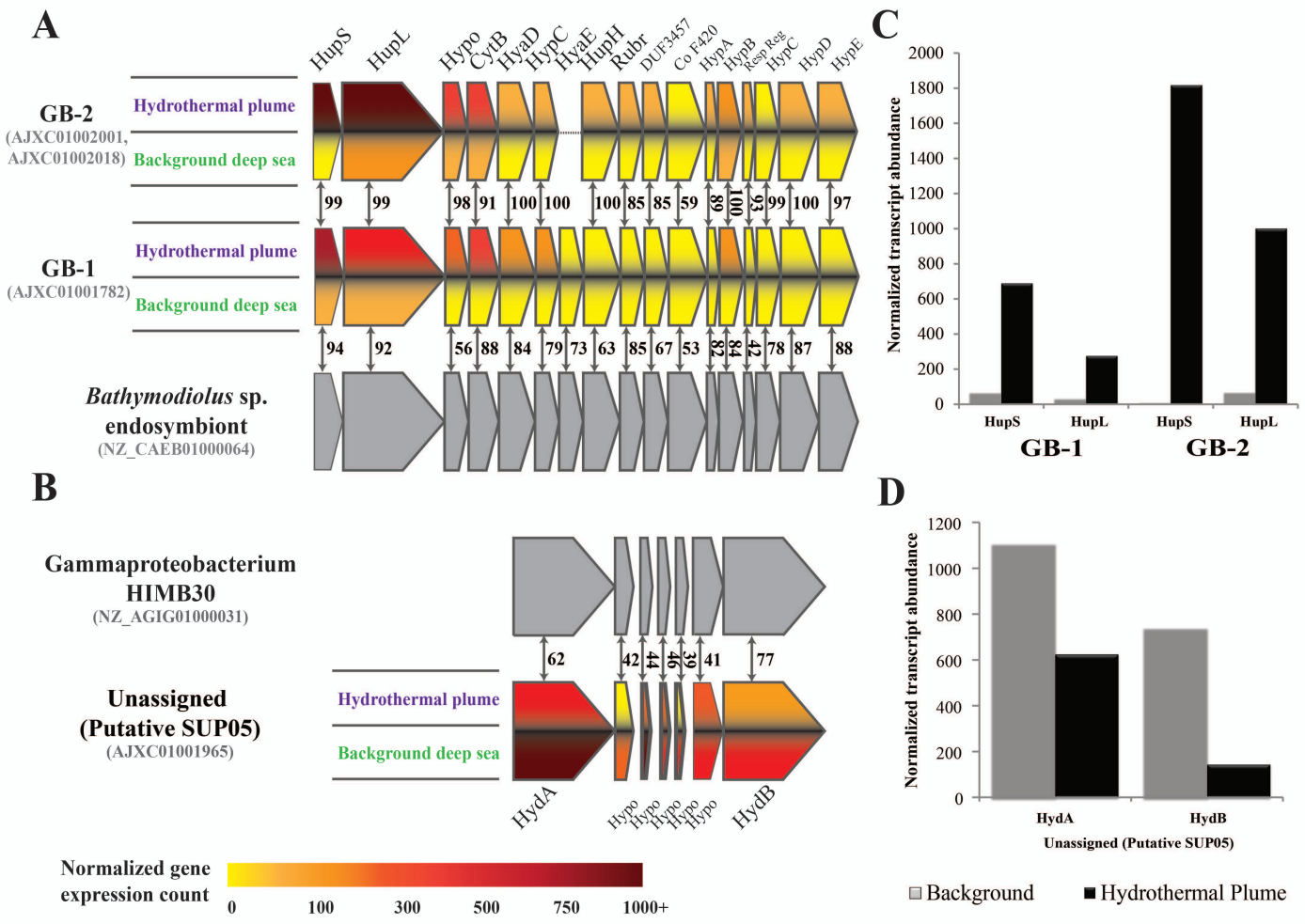
Figure 2. Map of pathways for sulfur oxidation by GB SUP05**.** Inset histograms depict the gene transcript abundance for individual genes in GB-1 and GB-2. Transcript abundance is normalized for gene length and total number of reads per dataset.

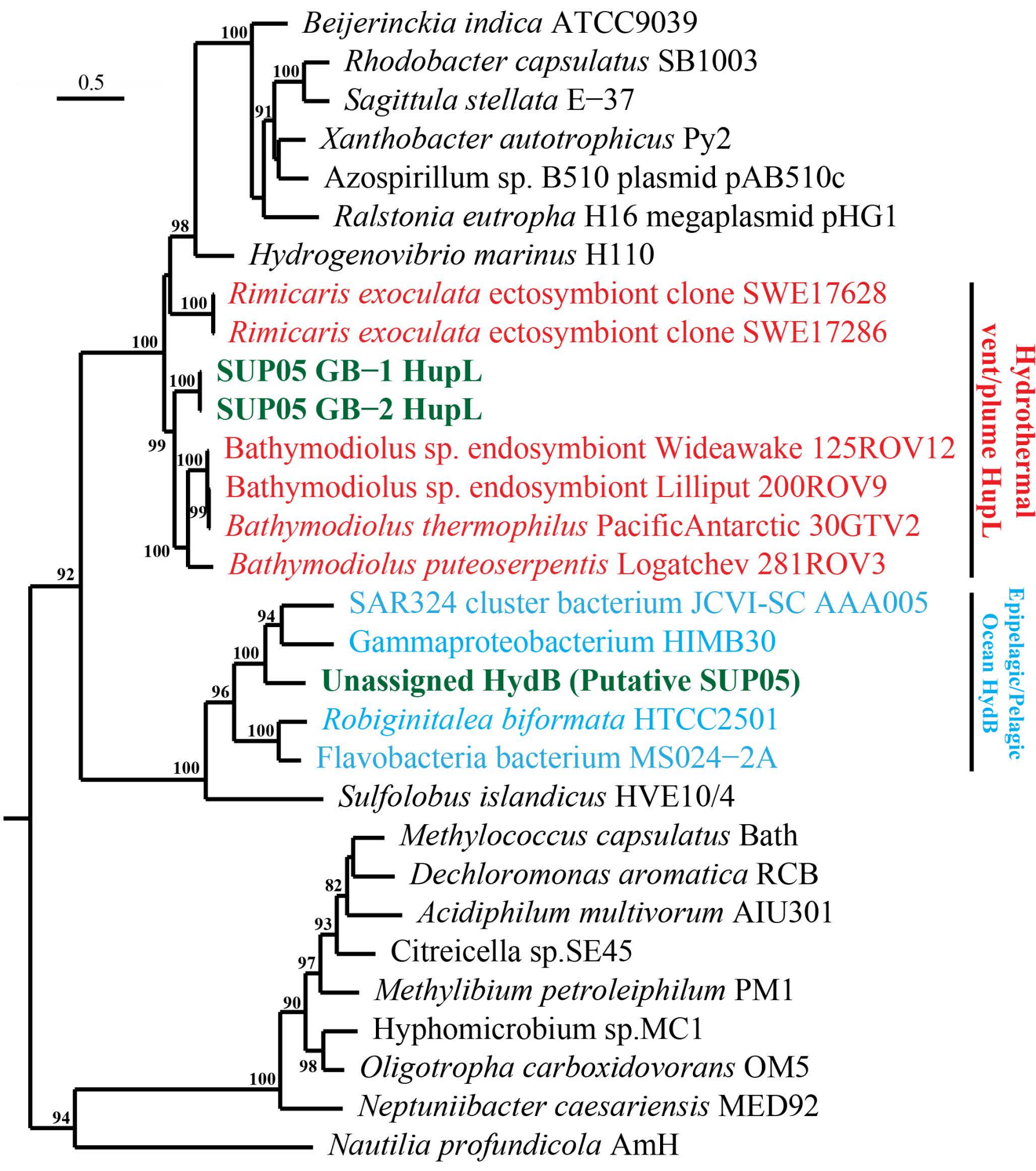
Figure 3. **A, B.** Organization and transcript abundance of GB-1 & 2 (A) and putative SUP05 (B) hydrogenase genes and comparison to closely related sequences from Genbank. Genes are colored according to normalized transcript abundance in plume and background. Arrows indicate shared genes and percent amino acid identity between predicted proteins. Dotted line in GB-2 indicates separation of contigs. **C, D**. Normalized transcript abundance for genes encoding small (HydA, HupS) and large subunits (HydB, HupL) of GB-1 & 2 (C) and putative SUP05 (D) hydrogenases.

Figure 4. Phylogeny of group 1 membrane bound Ni-Fe hydrogenase large subunit inferred with maximum likelihood. Bootstrap values greater than 80 are shown. Sequences in green are from Guaymas Basin, sequences in red are hydrothermal vent derived and sequences in blue are from the epipelagic ocean.


1
2
3
4
i
ii
iii
iv
v
vi
vii
viii
ix


Sulfide oxidation
900
800
700
600
500
400
300
200
100
0
GB-1 GB-2
sqr
HS-
sulfide
sqr
Background deep sea
Hydrothermal Plume
S0
Elemental sulfur
Sox pathway
S2O32-
Thiosulfate
Sulfur oxidation
1200
1000
800
600
400
200
0
GB-1 GB-2 GB-1 GB-2
dsrA dsrB
rdsr
Rhodanese sulfurtransferase
Thiosulfate oxidation
12000
10000
8000
6000
4000
2000
0
GB-1 GB-2
Rhodanese sulfurtransferase
SO32-
sulfite
aps
Sulfite oxidation
350
300
250
200
150
100
50
0
GB-1 GB-2 GB-1 GB-2 GB-1 GB-2
APS reductase,α
APS reductaseβ
sat
APS
adenosine
phosphosulfate
sat
Thiosulfate oxidation
4000
3000
2000
1000
0
GB-1 GB-2 GB-1 GB-2 GB-1 GB-2 GB-1 GB-2 GB-1 GB-2
SoxA SoxZ SoxY SoxX SoxB
Sox pathway
SO42-
sulfate


A
HupS
HupL
Hypo
CytB
HyaD
HypC
HyaE
HupH
Rubr
DUF3457
Co F420
HypA
HypB
Resp Reg
HypC
HypD
HypE
GB-2
(AJXC01002001, AJXC01002018)
Hydrothermal plume
Background deep sea
99 99 98 91 100 100 100 85 85 59 89 100 93 99 100 97
GB-1
(AJXC01001782)
Hydrothermal plume
Background deep sea
94 92 56 88 84 79 73 63 85 67 53 82 84 42 78 87 88
Bathymodiolus sp. endosymbiont
(NZ_CAEB01000064)
C
Normalized transcript abundance
2000
1800
1600
1400
1200
1000
800
600
400
200
0
HupS
HupL
HupS
HupL
GB-1
GB-2
B
Gammaproteobacterium HIMB30
(NZ_AGIG01000031)
Unassigned (Putative SUP05)
(AJXC01001965)
62 42 44 46 39 41 77
Hydrothermal plume
Background deep sea
HydA
Hypo
Hypo
Hypo
Hypo
Hypo
HydB
Normalized gene expression count
0 100 300 500 750 1000+
D
Normalized transcript abundance
1200
1000
800
600
400
200
0
HydA
HydB
Unassigned (Putative SUP05)
Background
Hydrothermal Plume


0.5
Beijerinckia indica ATCC9039
Rhodobacter capsulatus SB1003
Sagittula stellata E−37
Xanthobacter autotrophicus Py2
Azospirillum sp. B510 plasmid pAB510c
Ralstonia eutropha H16 megaplasmid pHG1
Hydrogenovibrio marinus H110
Rimicaris exoculata ectosymbiont clone SWE17628
Rimicaris exoculata ectosymbiont clone SWE17286
SUP05 GB−1 HupL
SUP05 GB−2 HupL
Bathymodiolus sp. endosymbiont Wideawake 125ROV12
Bathymodiolus sp. endosymbiont Lilliput 200ROV9
Bathymodiolus thermophilus PacificAntarctic 30GTV2
Bathymodiolus puteoserpentis Logatchev 281ROV3
Hydrothermal vent/plume HupL
SAR324 cluster bacterium JCVI-SC AAA005
Gammaproteobacterium HIMB30
Unassigned HydB (Putative SUP05)
Robiginitalea biformata HTCC2501
Flavobacteria bacterium MS024−2A
Epipelagic/Pelagic Ocean HydB
Sulfolobus islandicus HVE10/4
Methylococcus capsulatus Bath
Dechloromonas aromatica RCB
Acidiphilum multivorum AIU301
Citreicella sp.SE45
Methylibium petroleiphilum PM1
Hyphomicrobium sp.MC1
Oligotropha carboxidovorans OM5
Neptuniibacter caesariensis MED92
Nautilia profundicola AmH