Metabolic flexibility of enigmatic Sar324 revealed through metagenomics and metatranscriptomics

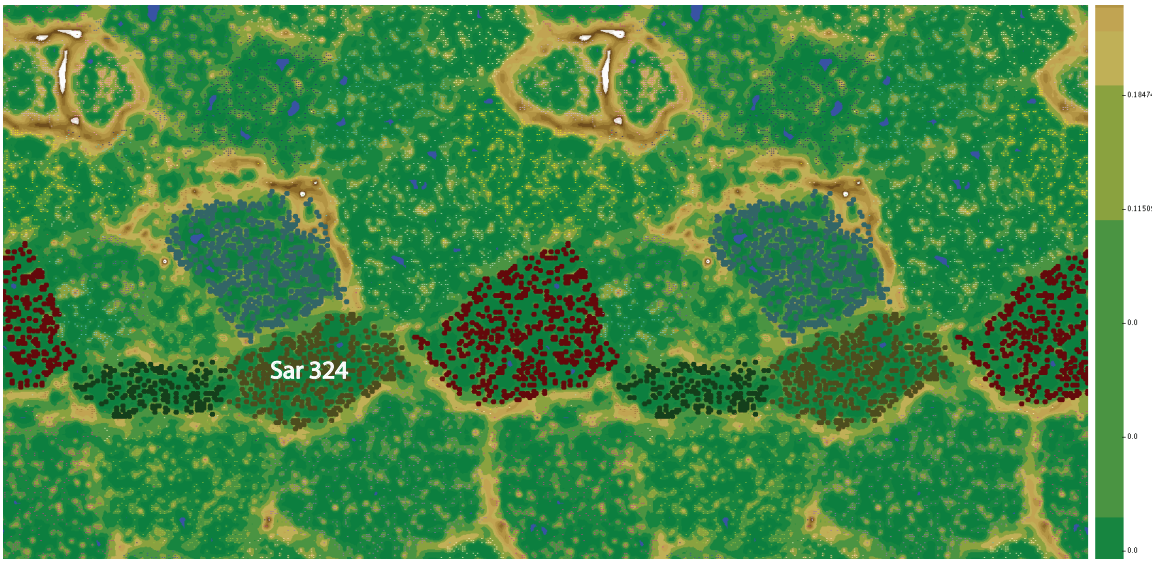
Short Title: Disentangling the ecophysiological role of Sar324

Cody S. Sheik[1], Sunit Jain[1] and Gregory J. Dick[1,2,3]

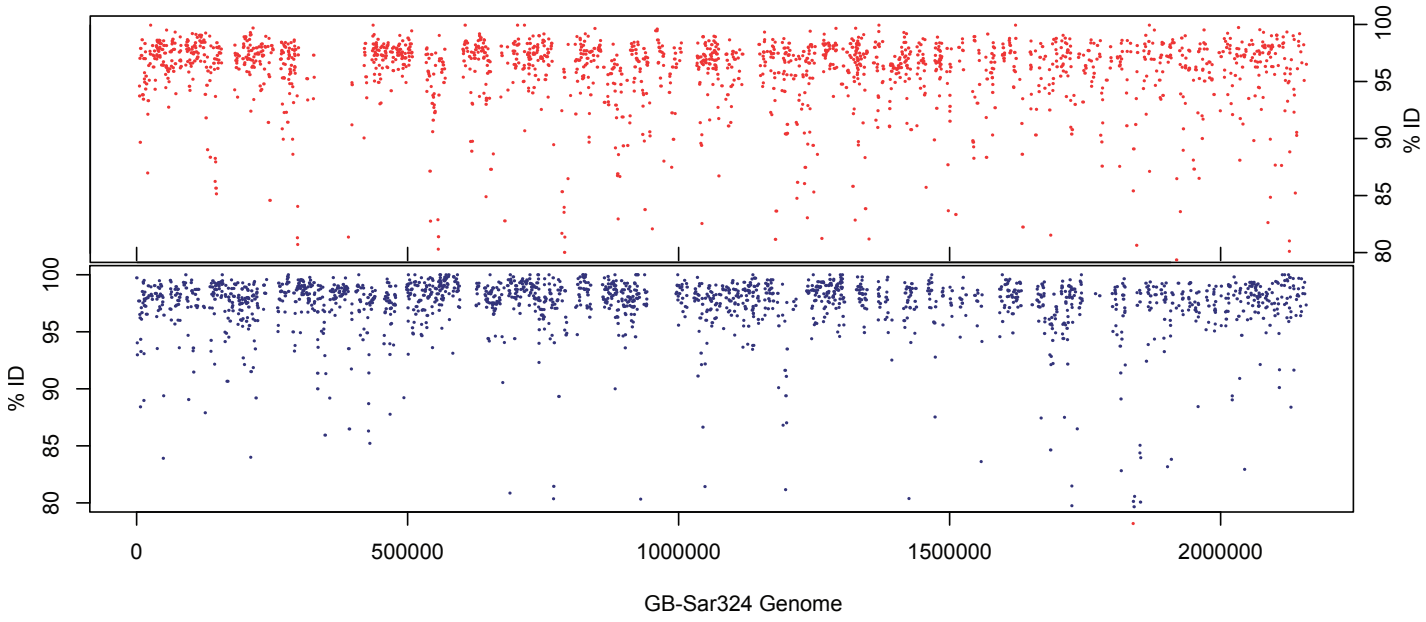
[1]Department of Earth and Environmental Sciences
[2]Center for Computational Medicine and Bioinformatics
[3]Department of Ecology and Evolutionary Biology
University of Michigan, Ann Arbor, MI, 48109, USA

Correspondence: Greg Dick, Department of Earth & Environmental Sciences, The University of Michigan, 1100 N. University Ave., 2534 CC Little Bldg, Ann Arbor MI, 48109-1005, USA. E-mail: gdick@umich.edu

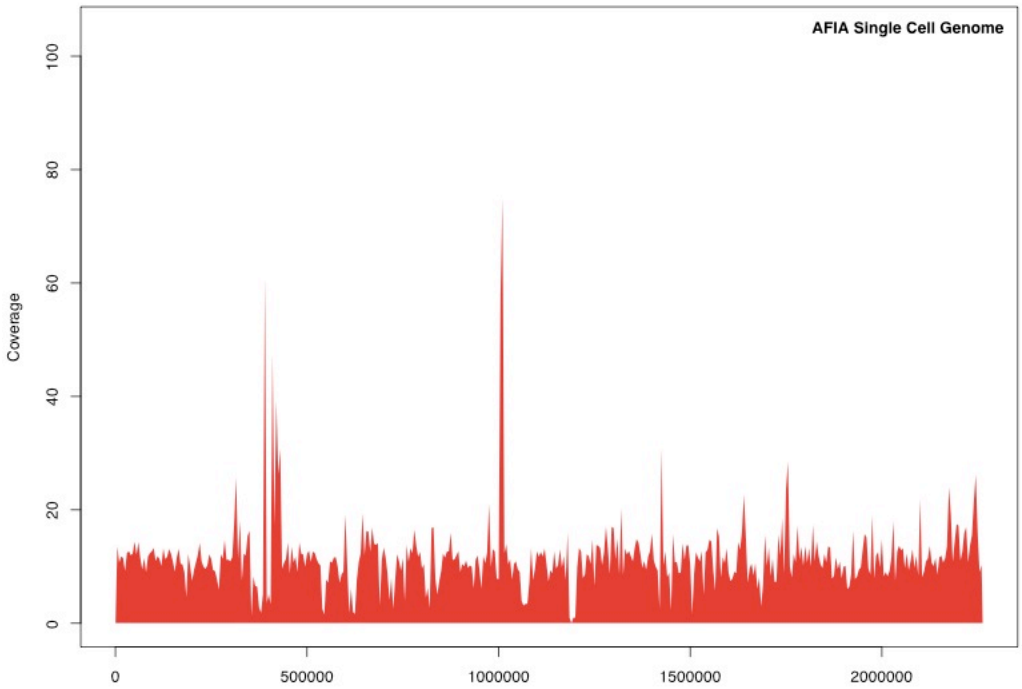
SI Figure 1. Emergent Self Organizing Map based on tetranucleotide frequencies from metagenomic contigs greater than 2500 bp. Each data point represents a genomic fragment (2500-5000 bp) and the background topology represents the Euclidean distance between tetranucleotide frequency profiles. Thus genomic bins are delineated by ridges. The GB-Sar324 bin is identified, and other major genomic bins are colored: Blue =MG1 Archaea, Dark Red= SUP05 and black = methylotrophs.

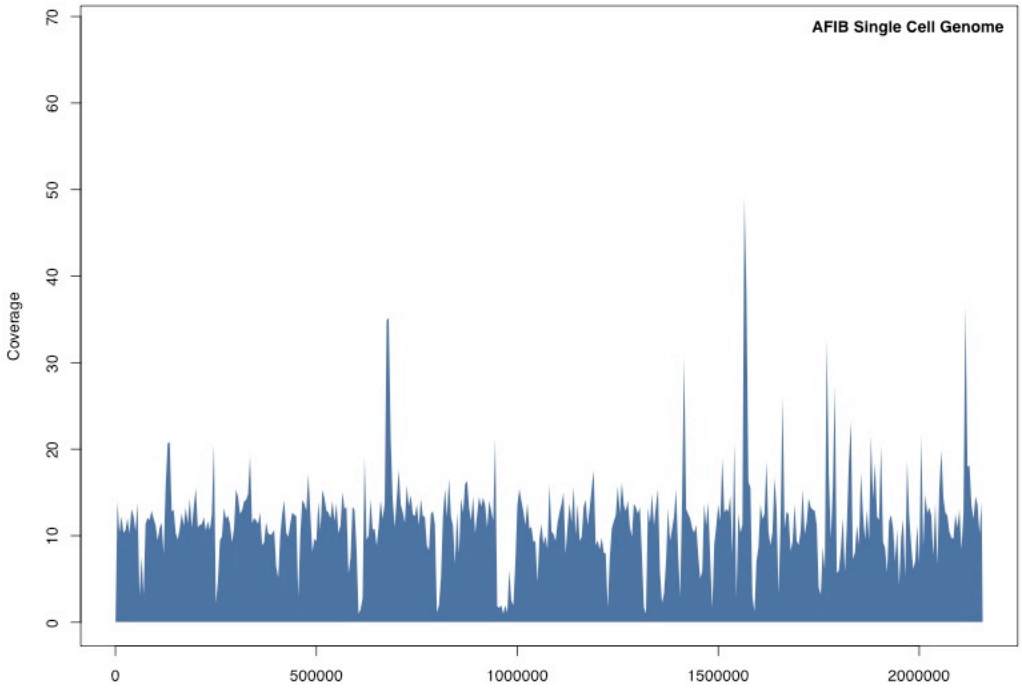


SI Figure 2. Recruitment of genes from published single cell Sar324 genomes to GB-Sar324 Metagenome Bin. Recruitment was performed using BLASTn with high stringency cutoffs (e-value cutoff 10^{-10}, minimum length= 400 bases). Blue dots represent AFIB (AFIB00000000) and red dots represent AFIA (AFIA00000000) single cell genomes. Blue and red lines represent the average nucleotide similarity to GB-Sar324.

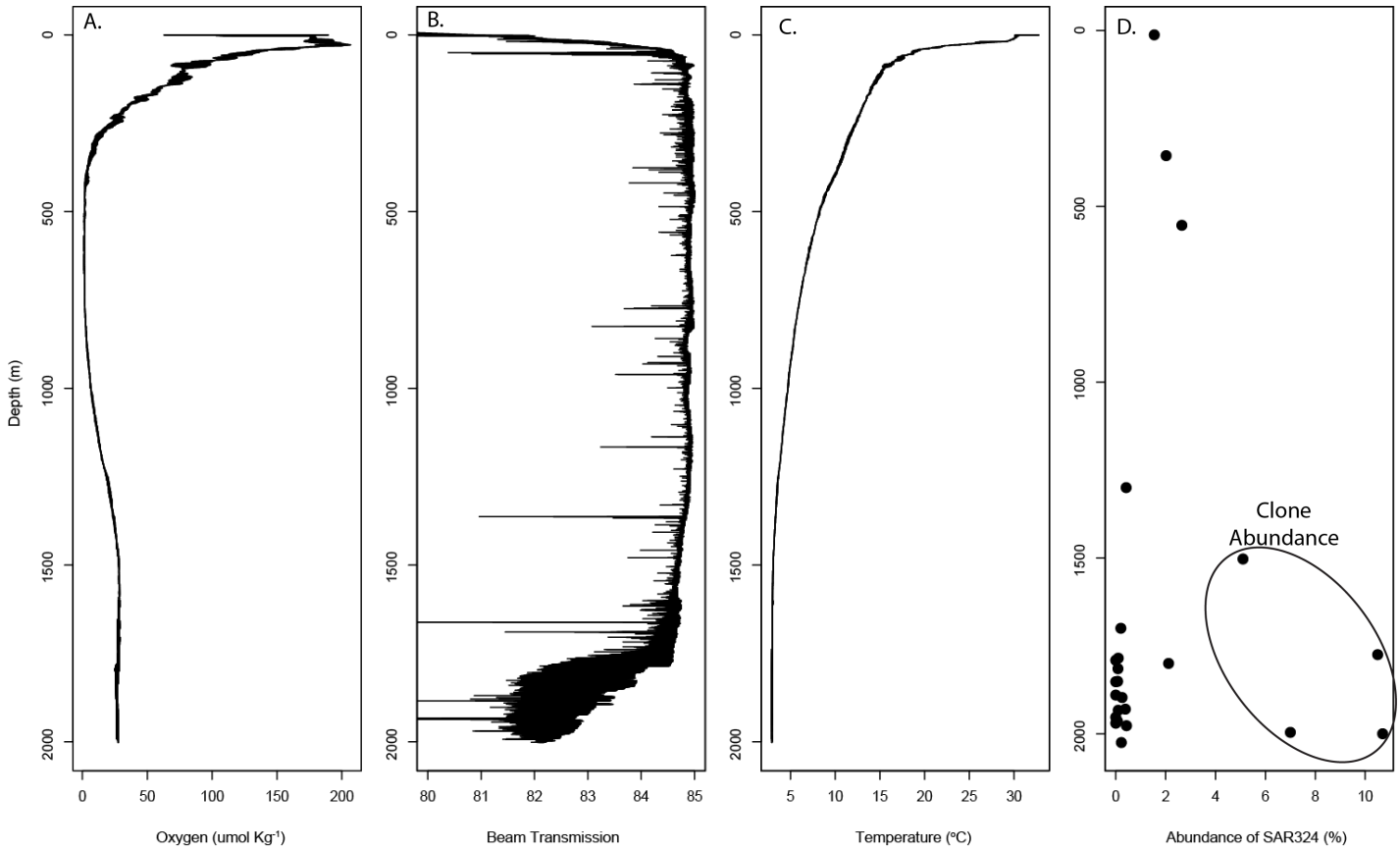


SI Figure 3. Recruitment of Guaymas Basin 454 metagenomic reads to the dark ocean SAR324 single amplified genomes. Reads were recruited with BLASTn using a similarity of >90% and bit scores over 70. Coverage was calculated with a 100 base pair sliding window.

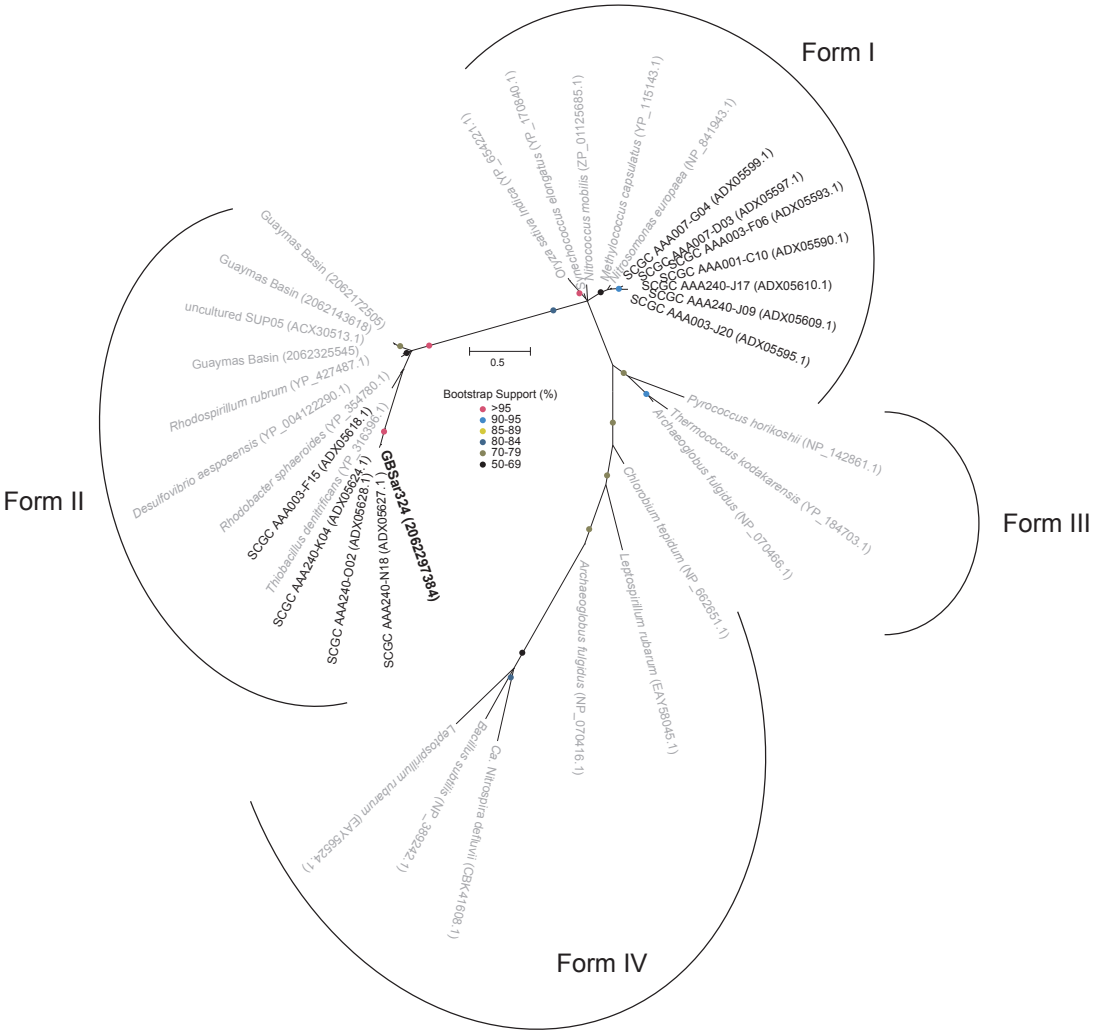




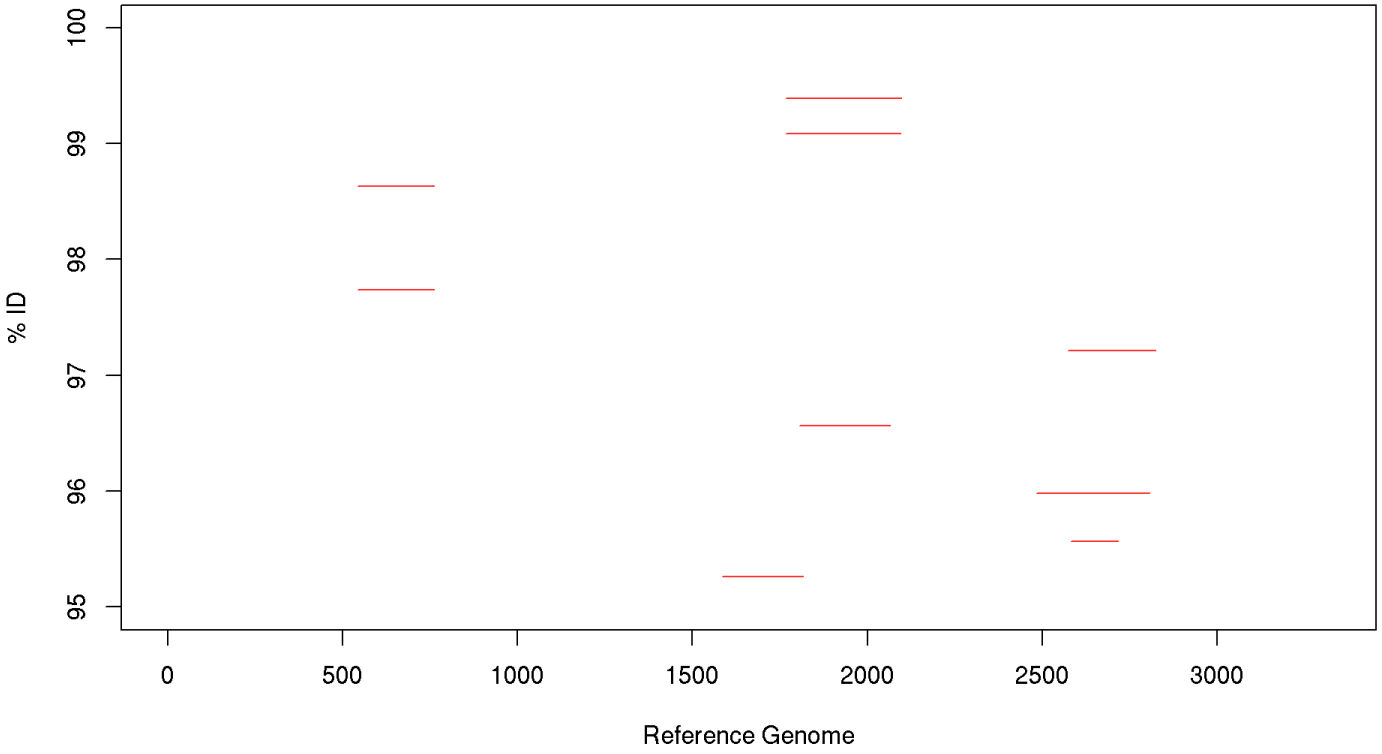
SI Figure 4. Representative CTD depth profile from Guaymas Basin. A) Oxygen concentration, B) Beam transmission, C) Temperature and D) Sar324 abundance from clone libraries (Dick and Tebo, 2010) and high throughput tagged ribosomal sequencing libraries (Anantharaman, 2012). Note that the clone library data is as a percentage of Bacteria, whereas the tag sequence data is as a percentage of Bacteria plus Archaea.



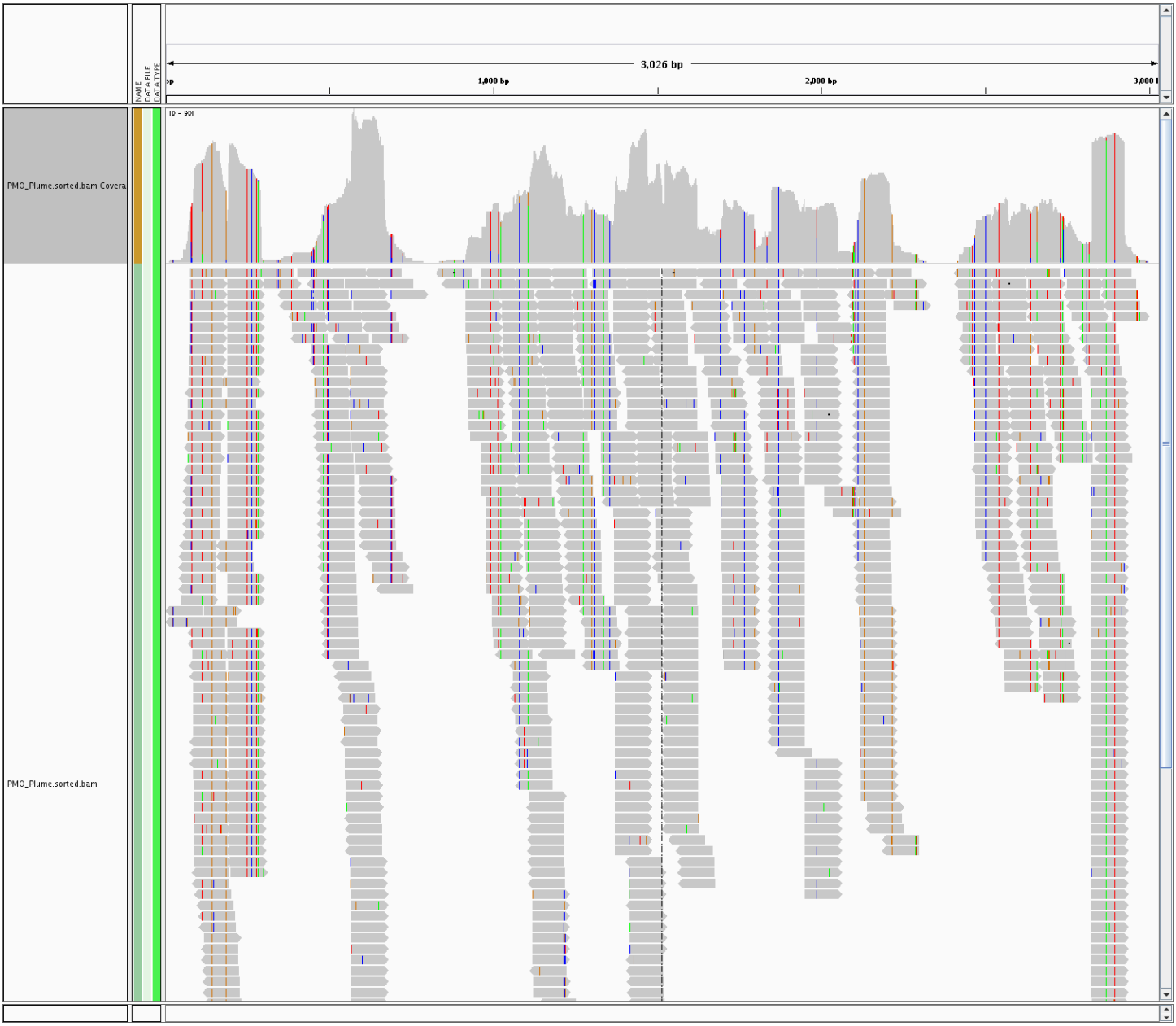
SI Figure 5. Phylogenetic inference of large subunit 1,5-ribulose bisphosphate genes (RuBisCO). Sar324 related sequences are highlighted in black and the GB-Sar324 RuBisCO gene is bolded.



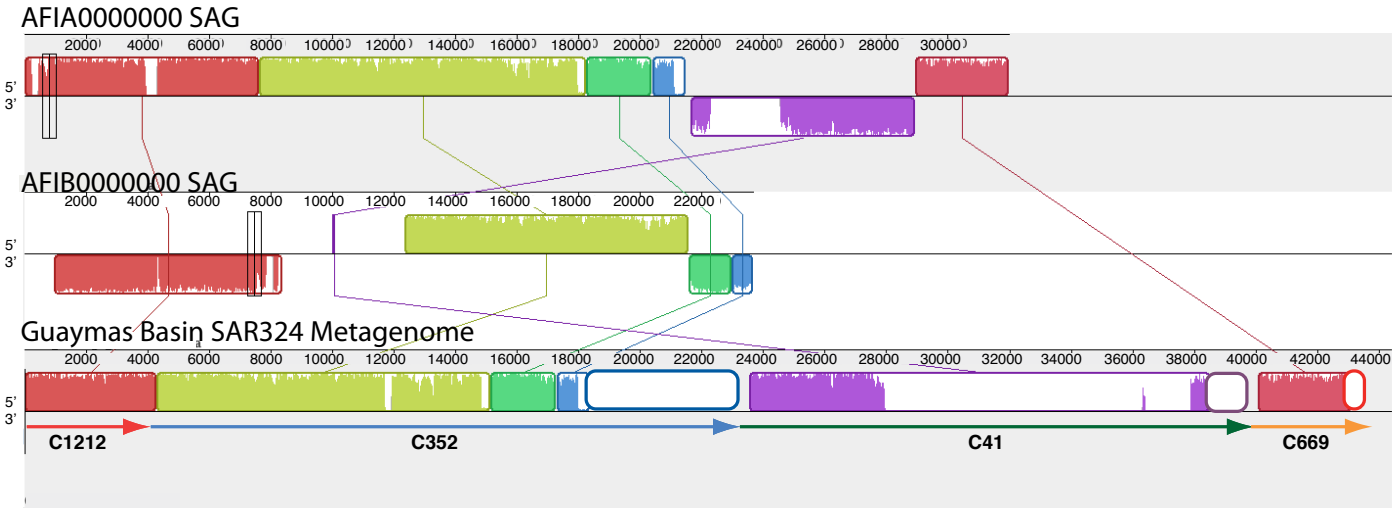
SI Figure 6. Recruitment of 454 metagenome reads to the phmoABC operon from SAR324 single amplified genome (AAA240-J09). Reads were recruited with BLASTn using a similarity of >90% and bit scores over 70.



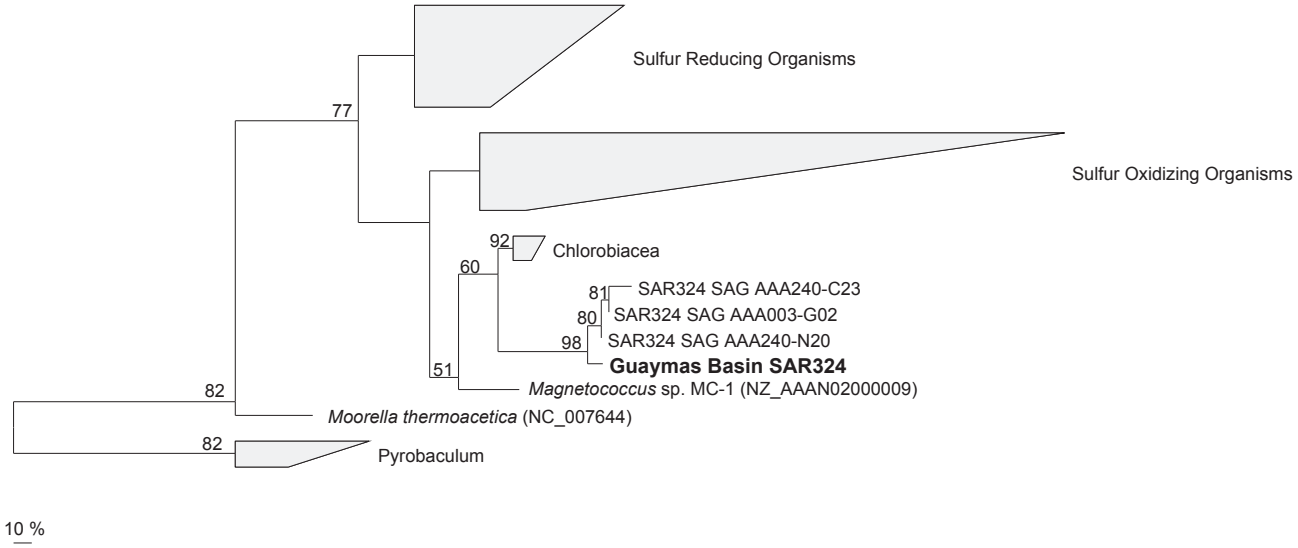
SI Figure 7. Recruitment of paired end illumina metatranscriptome reads mapped to the phmoABC operon from SAR324 single amplified genome (AAA240-J09). Reads were mapped with BWA and alignments were visualized with IGV (See methods section). The order of genes for the Phmo are CAB, which can clearly be seen through mapping.



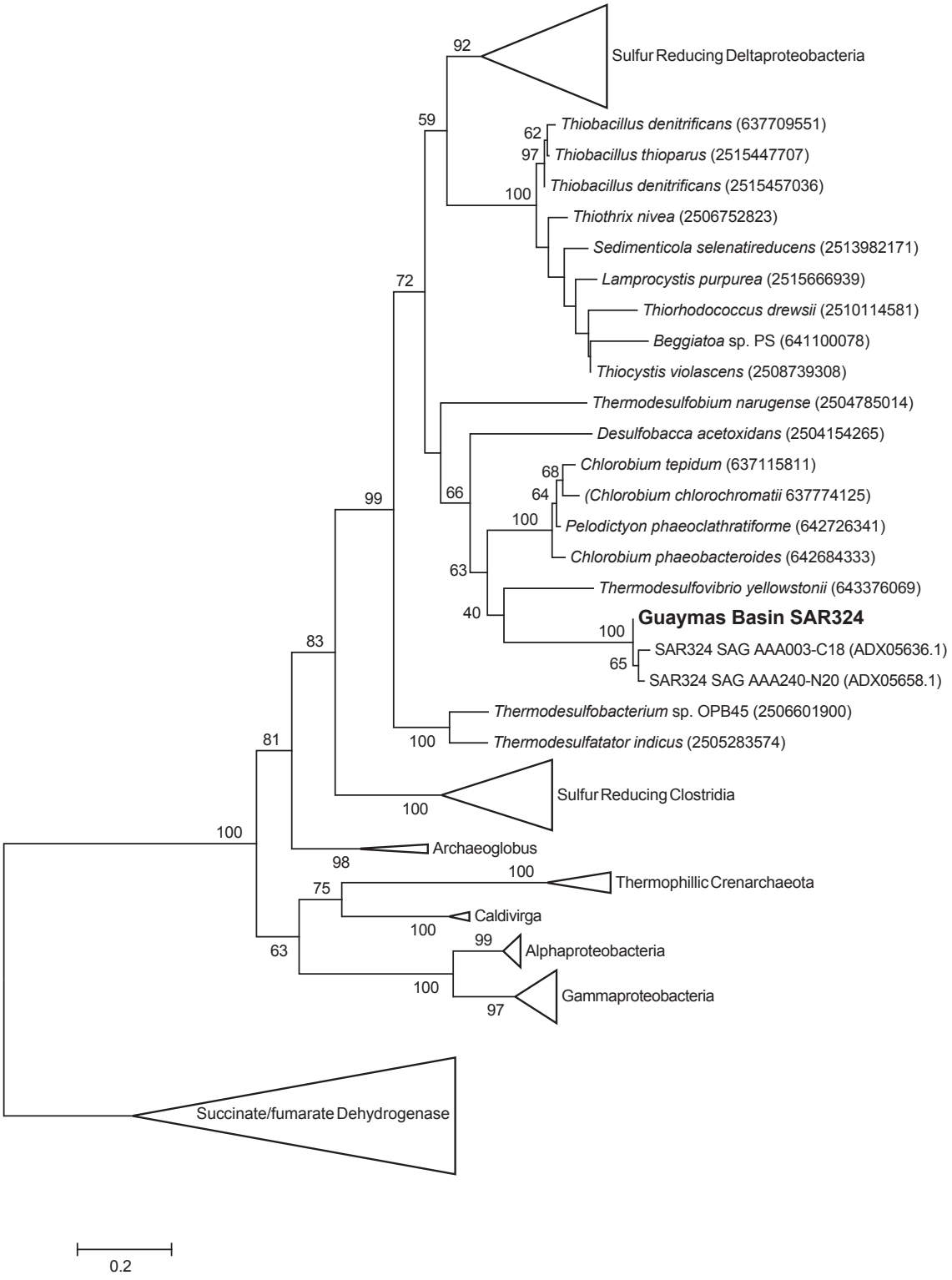
SI Figure 8. Depiction of contig synteny in GB-Sar324 and single cell genome isolates. Contigs all contain sulfur cycling genes, the expression profiles of representative genes are shown in Figure 3. These five contigs from GB-SAR324 are present in as single contig in the AFIA SAG (AAA240-J09), thus is used as a reference. Colors represent contigs from GB-SAR324, while inside contig boxes lines represent similarity between GB-SAR324 and SAGs



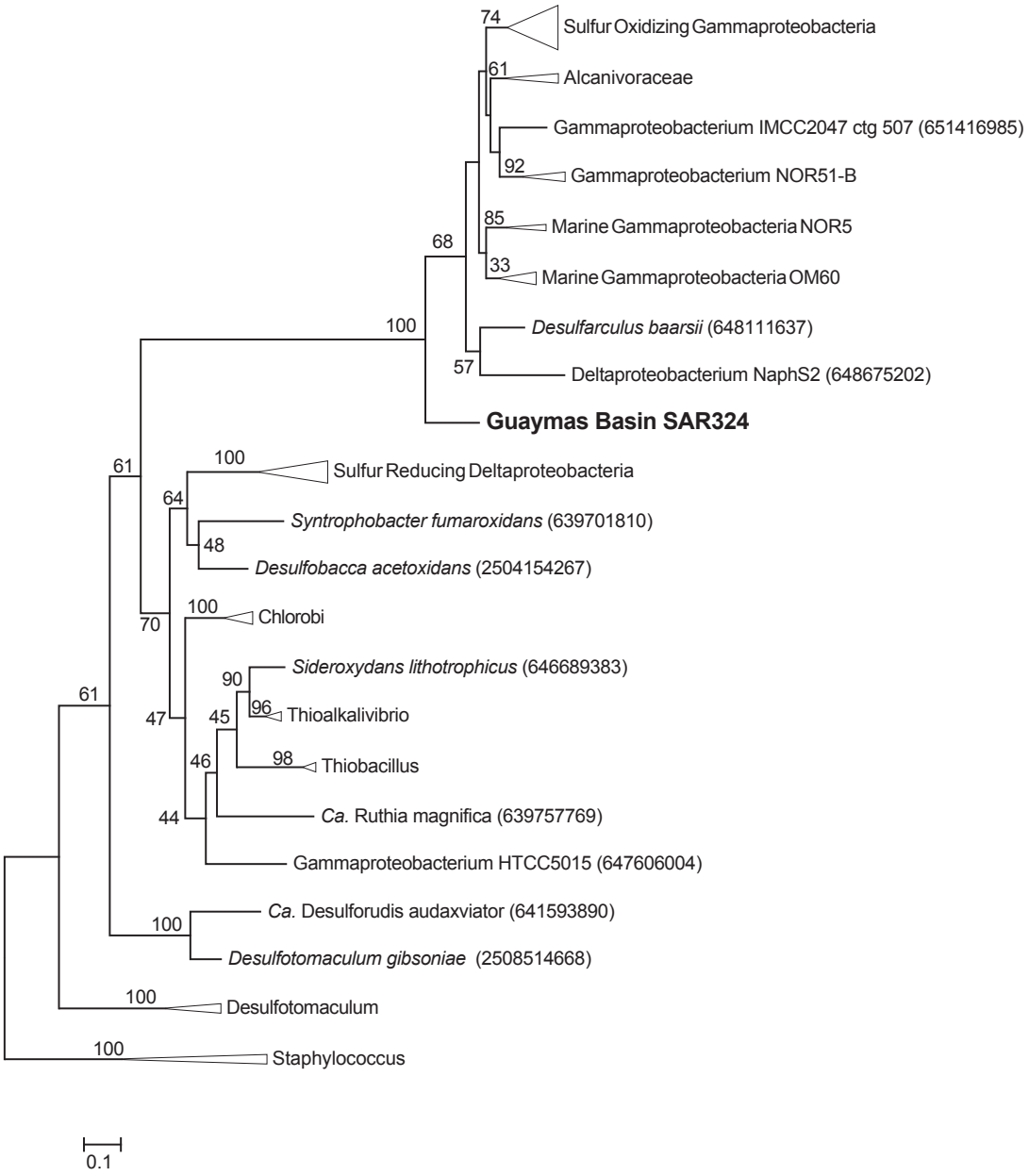
SI Figure 9. Phylogeny of the *dsr*A gene from SAR324 in relation to sulfur oxidizing and sulfur reducing microorganisms.



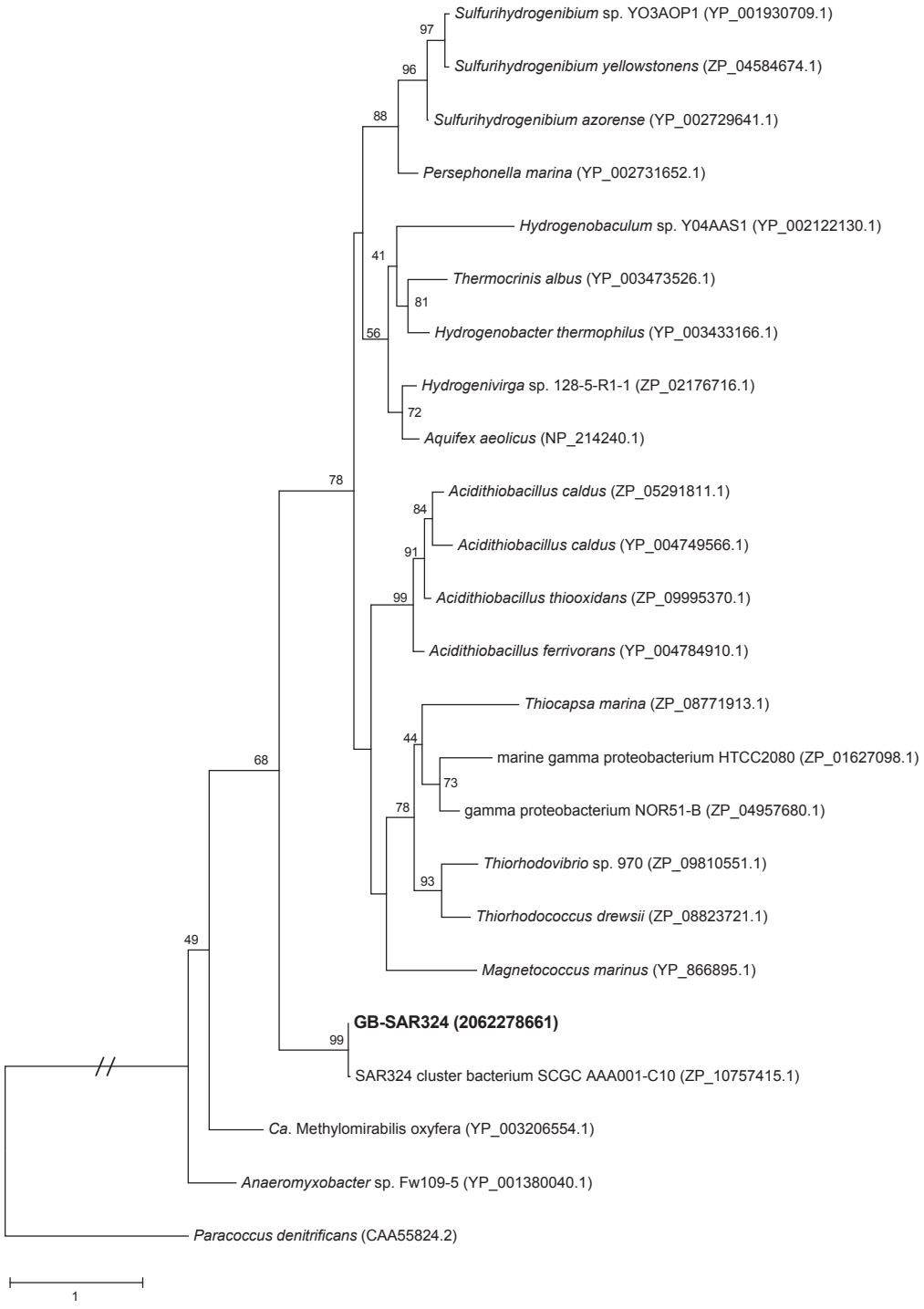
SI Figure 10. Phylogeny of the *apr*A gene from SAR324 to sulfur oxidizing and sulfur reducing microorganisms.



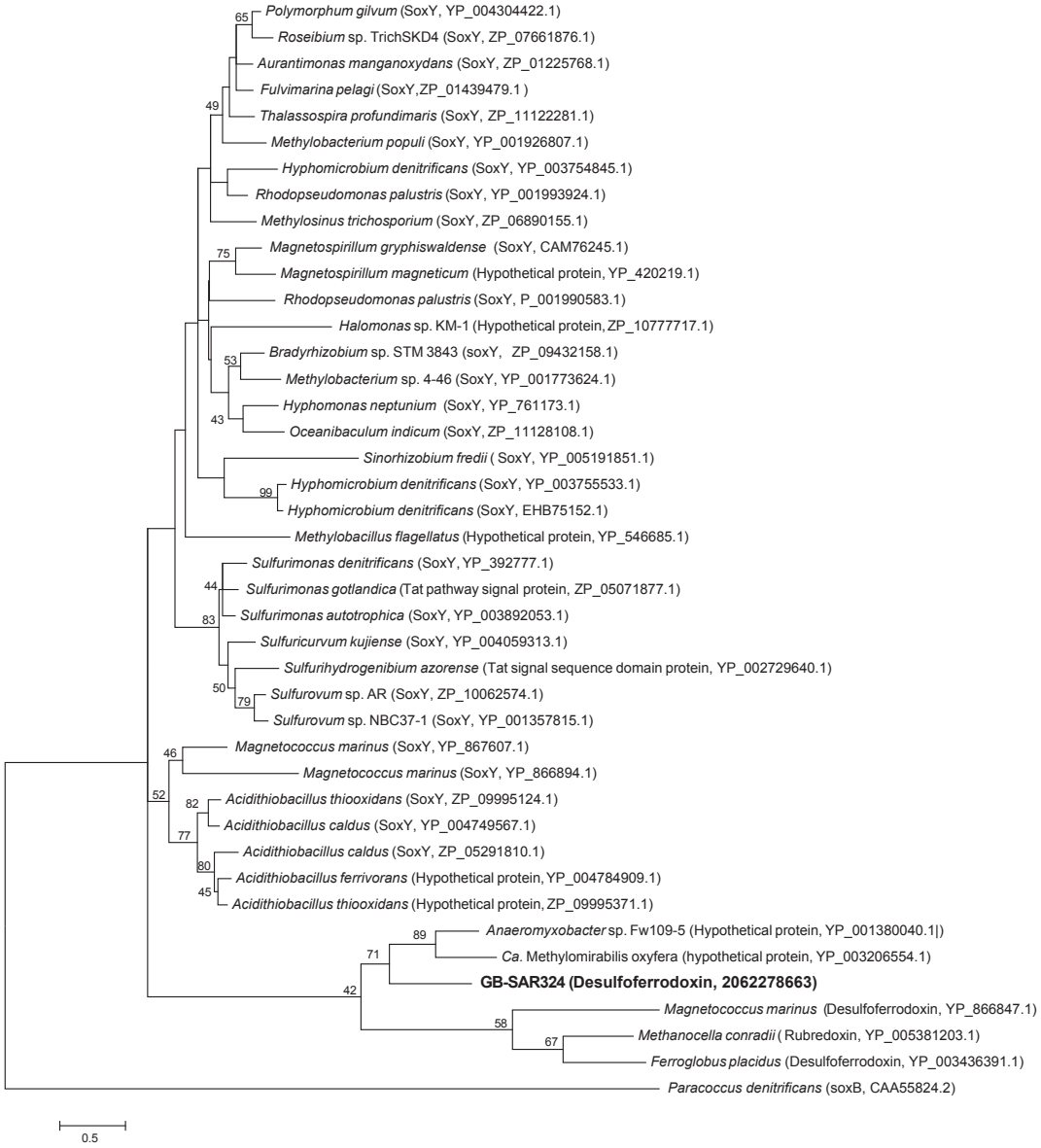
SI Figure 11. Phylogeny of the SAT gene from GB-SAR324 to sulfur oxidizing and sulfur reducing microorganisms.



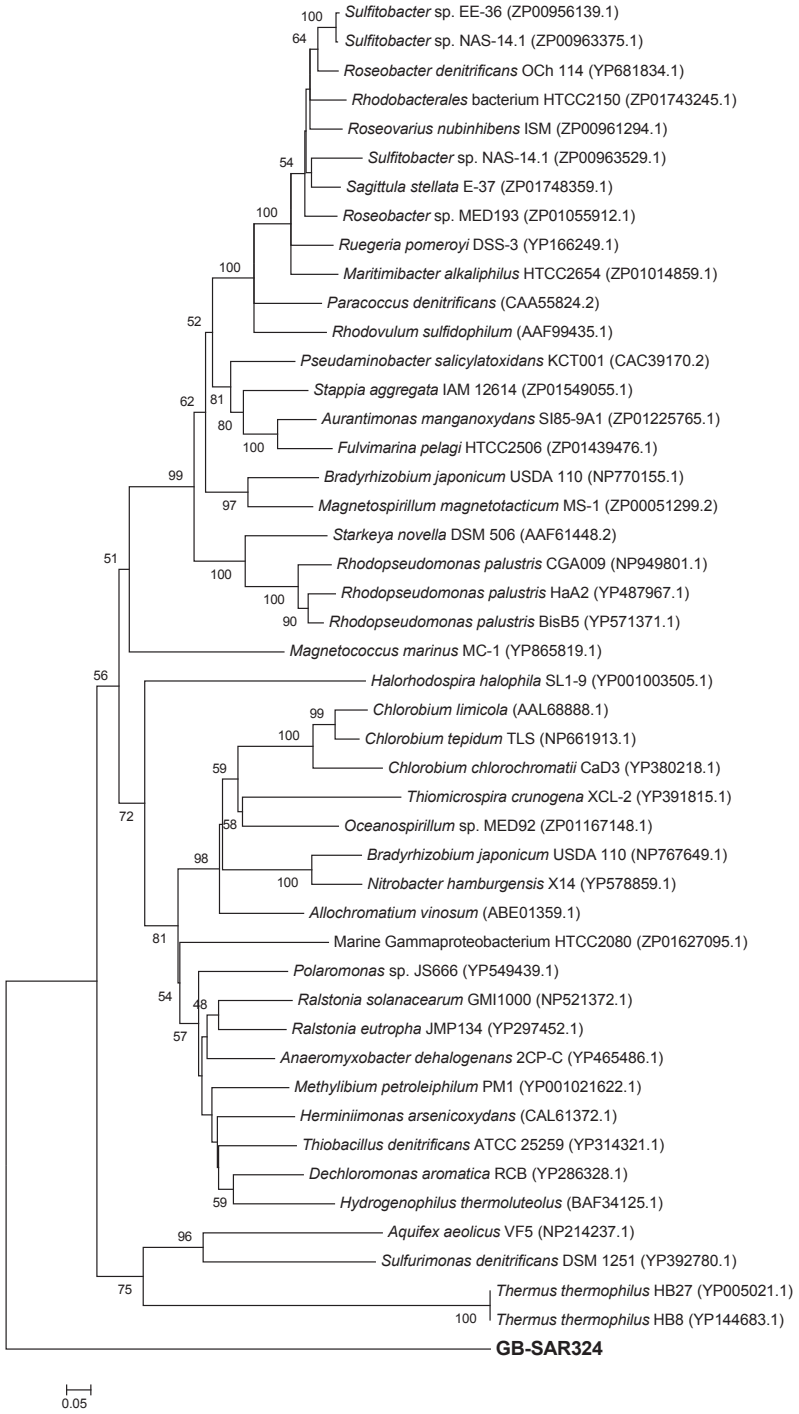
SI Figure 12. Phylogeny of the SoxZ-like gene from SAR324 to sulfur oxidizing microorganisms.



SI Figure 13. Phylogeny of the SoxY-like gene from GB-SAR324 to sulfur oxidizing microorganisms and phylogenetically similar Desulfoferrodoxin genes.

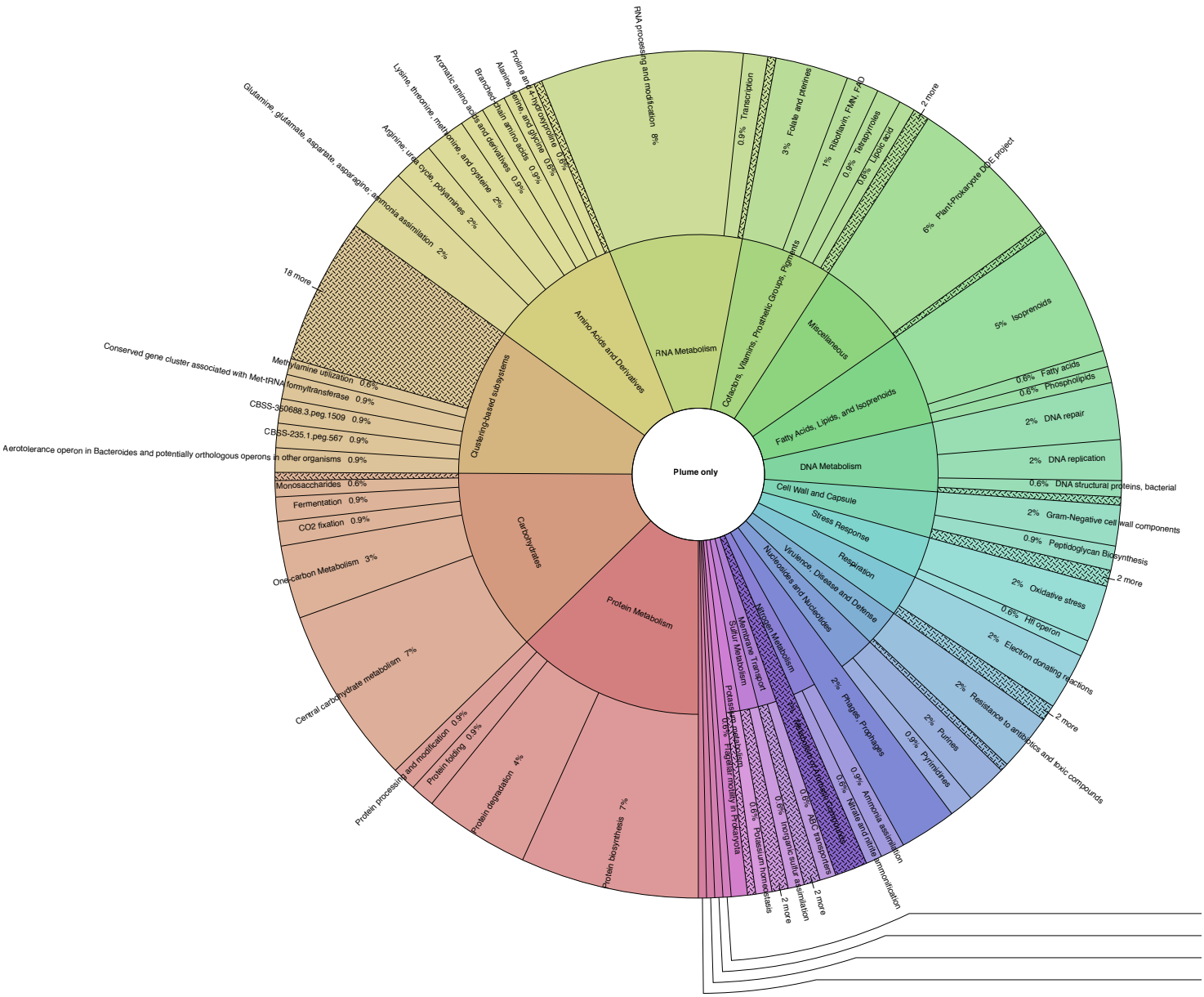


SI Figure 14. Phylogeny of the SoxB-like gene from GB-SAR324 to sulfur oxidizing microorganisms.



SI Figure 15. Genes detected only in the background cDNA metatranscriptome. Percentages are calculated as fraction of genes found only in the background cDNA libraries.

SI Figure 16. 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.



SI Figure 17. 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 color coded for reference in closest relative genomes. * 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.