Supplemental material for:

Metatranscriptomic analysis of diminutive *Thiomargarita*-like bacteria ("*Candidatus* Thiopilula spp.") from abyssal cold seeps of the Barbados Accretionary Prism

Daniel S. Jones, Beverly E. Flood, Jake V. Bailey

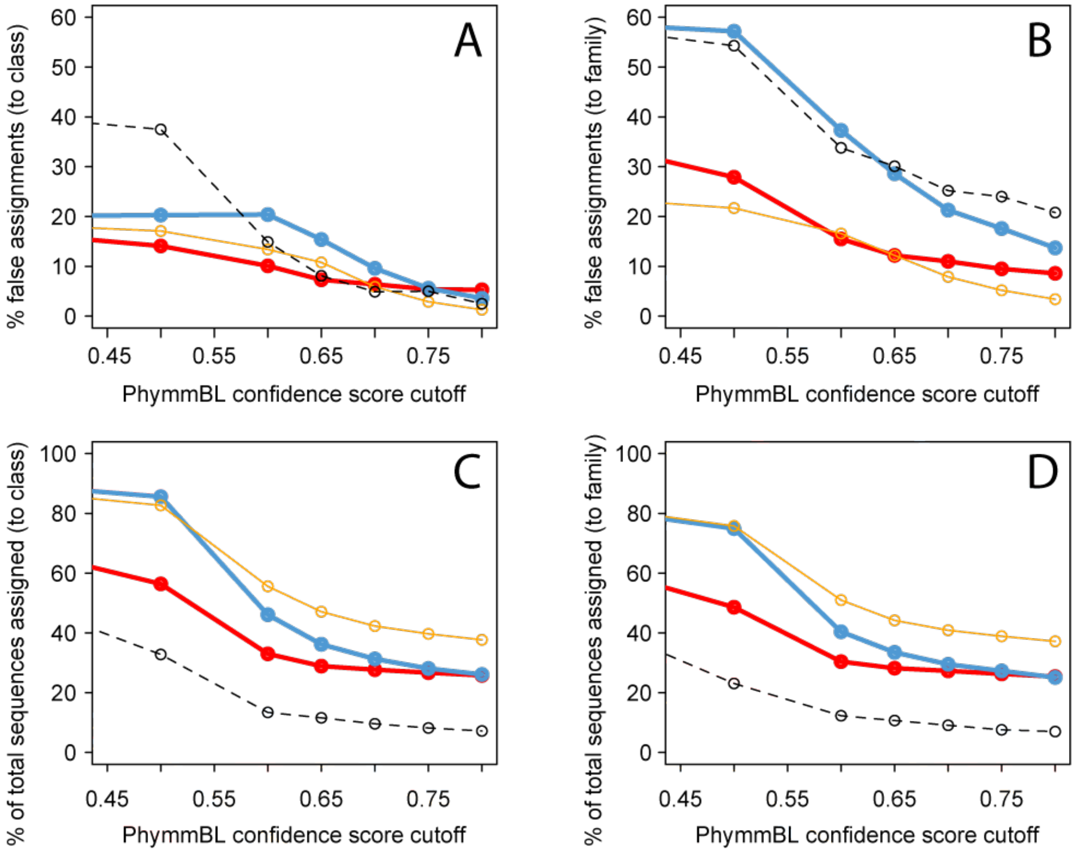

FIG S1. Classification of simulated genome fragments. Four draft genomes that were not present in our PhymmBL database were each fragmented into 1 kilobase-pair segments *in silico*, and each segment classified with PhymmBL. Panels A and B show the percentage of false assignments at the class and family taxonomic levels, respectively. Panels C and D show the percentage of total sequences that were assigned at the class and family levels, respectively. The three genomes in solid lines (*Photobacterium phosphoreum*, thick blue; *Acidithiobacillus thiooxidans*, thick red; and *Sulfurospirillum multivorans,* thin orange) had one other representative of their genus present in the PhymmBL database, while the black dashed line represents a genome (*Methanomicrobium mobile*) that was only represented at the family level. At the 0.65 confidence level, we observed >85% and >70% accurate assignments at the class and family levels, respectively, although <50% of total sequences could be assigned with this cutoff.

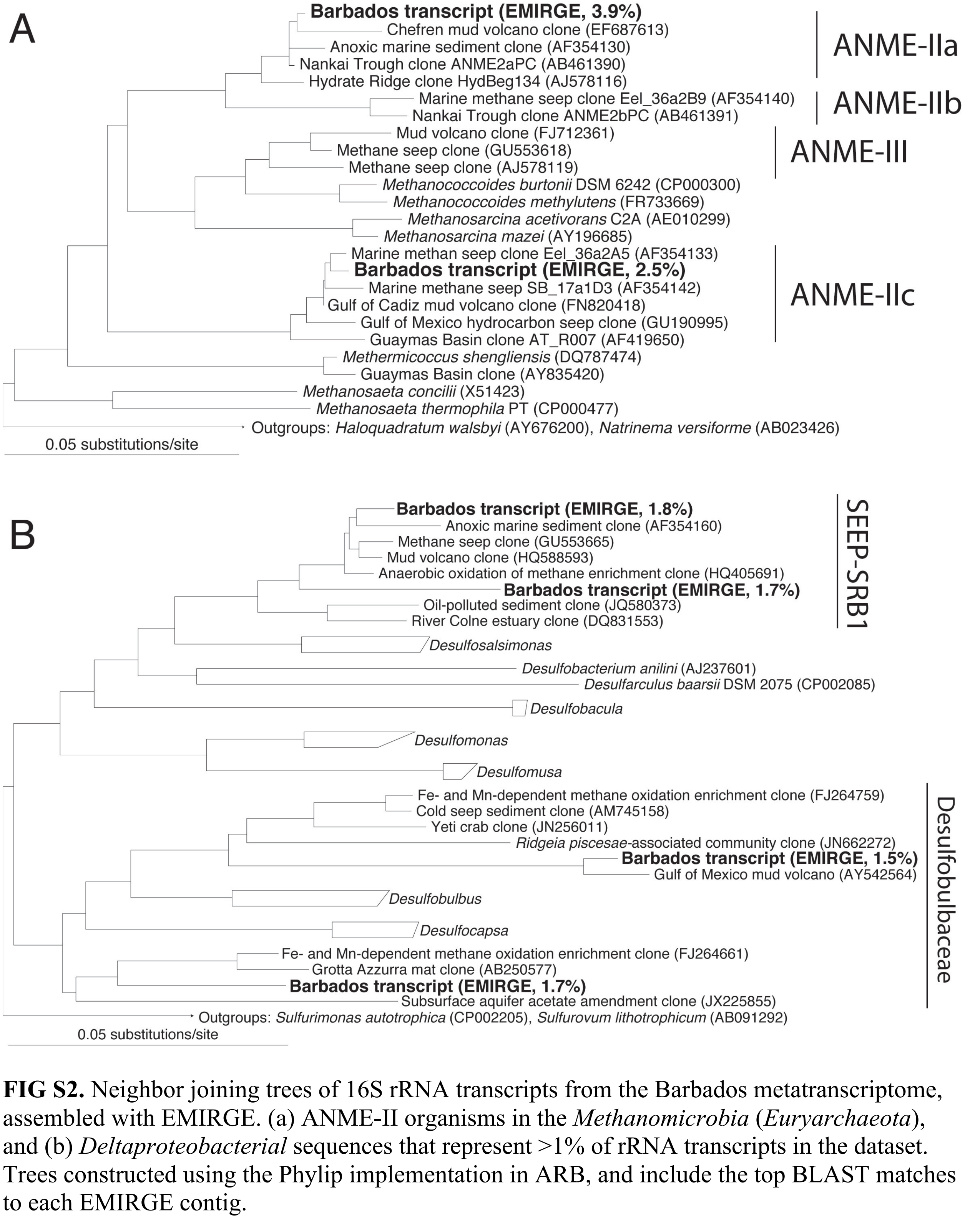

FIG S2. Neighbor joining trees of 16S rRNA transcripts from the Barbados metatranscriptome, assembled with EMIRGE. (a) ANME-II organisms in the *Methanomicrobia* (*Euryarchaeota*), and (b) *Deltaproteobacterial* sequences that represent >1% of rRNA transcripts in the dataset. Trees constructed using the Phylip implementation in ARB, and include the top BLAST matches to each EMIRGE contig.

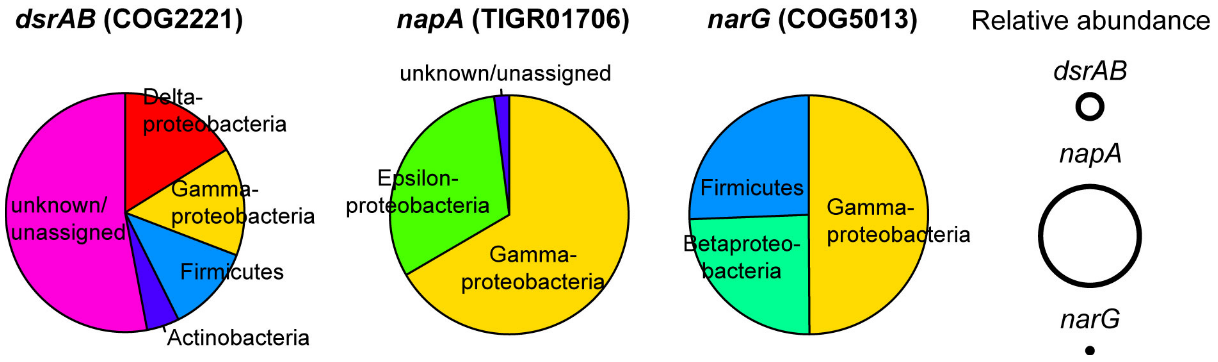

FIG S3. Taxonomic affiliation of genes for nitrate and sulfate reduction in the metatranscriptome: dissimilatory sulfite reductase *dsrAB*, periplasmic nitrate reductase *napA*, and membrane-bound nitrate reductase *narG*. Genes were identified in the RAMMCAP annotation using COG and TIGRfam assignments. (*napA* does not have a specific COG category but it is represented by TIGRfam TIGR01706.) Based on coverage, *napA* is 23x and 4x times more abundant than *narG* and *dsrAB* in the metatranscriptome, respectively (relative abundance, right).

Supplemental Table S1. Universal marker COGs used for taxonomic classification, according to Ciccarelli et al. (1)

COG ID	COG description
COG0012	Predicted GTPase, probable translation factor
COG0016	Phenylalanyl-tRNA synthetase alpha subunit
COG0048	Ribosomal protein S12
COG0049	Ribosomal protein S7
COG0052	Ribosomal protein S2
COG0080	Ribosomal protein L11
COG0081	Ribosomal protein L1
COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit
COG0087	Ribosomal protein L3
COG0091	Ribosomal protein L22
COG0092	Ribosomal protein S3
COG0093	Ribosomal protein L14
COG0094	Ribosomal protein L5
COG0096	Ribosomal protein S8
COG0097	Ribosomal protein L6P/L9E
COG0098	Ribosomal protein S5
COG0099	Ribosomal protein S13
COG0100	Ribosomal protein S11
COG0102	Ribosomal protein L13
COG0103	Ribosomal protein S9
COG0172	Seryl-tRNA synthetase
COG0184	Ribosomal protein S15P/S13E
COG0186	Ribosomal protein S17
COG0197	Ribosomal protein L16/L10E
COG0200	Ribosomal protein L15
COG0201	Preprotein translocase subunit SecY
COG0202	DNA-directed RNA polymerase, alpha subunit/40 kD subunit
COG0256	Ribosomal protein L18
COG0495	Leucyl-tRNA synthetase
COG0522	Ribosomal protein S4 and related proteins
COG0533	Metal-dependent proteases with possible chaperone activity

SUPPLEMENTAL REFERENCES

1. **Ciccarelli FD, Doerks T, Von Mering C, Creevey CJ, Snel B, Bork P.** 2006. Toward automatic reconstruction of a highly resolved tree of life. Science **311:**1283-1287.