

Supplementary Figures

Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages

Kristopher Kieft[#], Zhichao Zhou[#], Rika E. Anderson, Alison Buchan, Barbara J. Campbell, Steven J. Hallam, Matthias Hess, Matthew B. Sullivan, David A. Walsh, Simon Roux, Karthik Anantharaman^{*}

[#]These authors contributed equally

^{*}Corresponding author

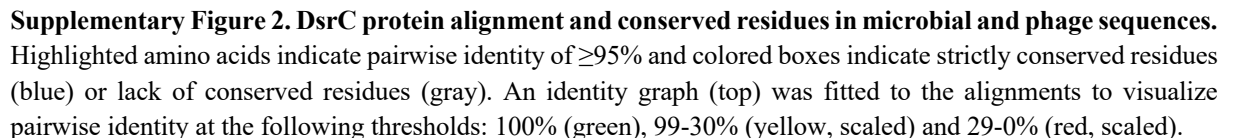
Candida, *maquilingensis* J167
Candida, *Hydrothermarchaeota*, *archaeae*, *g0JR_17*, *g0J2402035080*, *1001234*
Candida, *Hydrothermarchaeota*, *archaeae*, *g0JR2402035080*, *1001234*
Candida, *Rokubacteria*, *bacterium*, *RFCSPLOWO2_02_FULL_68_9*
COI_02_FULL, *Nitrosopiraeae_14*, *curated*
Chroococcum
Desulfobacter, *alkalithermophiles*
Moorella, *thermoaerobia*
Nitrosopiraeae, *bacterium*, *RFCSPLOWO2_02_FULL_39_10*
Pyrobaculum, *arsenatenum_1*
Pyrobaculum, *arsenatenum_2*
Abe, *10000269_1*
Ga0052187, *10001_184*
Ga0052187, *10002_1*
Ga0052187, *10007_1*
Ga0052187, *10045_82*
Ga0052187, *10045_83*
Ga0052187, *10048_80*
Ga0052187, *10048_81*
Ga0052187, *10081_115*
Ga0052187, *10083_108*
Ga0052187, *10086_102*
Ga0052187, *10089_127*
Ga0052187, *10090_49*
Ga0052187, *10091_49*
Ga0052187, *10092_52*
Ga0059848, *100117_1*
Ga0059848, *100137_11*
Ga0115000, *10010187_6*
Ga0163110, *10000569_26*
Ga018157, *1001787_1*
Ga0181581, *10002005_17*
Ga0181584, *10000368_52*
Ga0211543, *10000381_39*
g0J2404015324, *1001267_3*
KiloMoana, *1000167_70*
KiloMoana, *10003372_10*
KiloMoana, *10006689_38*
Mariner, *1000093_36*
Mariner, *1000093_37*
Mariner, *1000289_29*
TaHMoana, *1000070_25*
TaHMoana, *10020752_1*
TuMailla, *1001497_2*
Ga0181567, *1001775_23*
Ga0181568, *10002880_28*
Ga0181569, *10001662_26*
Ga0181571, *1001624_23*
Ga0181572, *10010653_23*

Caillivaudia, *maquilingensis* J167
Caillivaudia, *Hydrothermarchaeota*, *archaeae*, JRR_17_GJ2402035080, 1001234
Caillivaudia, *Hydrothermarchaeota*, *bacterium*, RIFCSPH4G02_02_FULL_68_19
Caillivaudia, *Rokubacteria*, *bacterium*, RIFCSPLOW02_02_FULL_68_19
COI, *COI*, *FULL*, *Nitrosopraeae*, J1_54, current
Chlorobium, *Chlorobium*
Desulfibacter, *alkalithermus*
Mosella, *thermarchaeota*
Nitrosibacter, *bacterium*, RIFCSPLOW02_02_FULL_39_10
Pyrobaculum, *arsenicatum*_1
Pyrobaculum, *arsenicatum*_2
Abe, 10000169_1
Abe, 10000200_59
GA0052187, 10001184
GA0052187, 100021_1
GA0052187, 100027_1
GA0052187, 100045,82
GA0052187, 100045,83
GA0052187, 100048,80
GA0052187, 100048,81
GA0052187, 100081,115
GA0052187, 100083,108
GA0052187, 100086,102
GA0052187, 100089,127
GA0052187, 100099,49
GA0052187, 100091,49
GA0052187, 100092,52
GA0098466, 1001117_6
GA0098466, 1001378_11
Kil155002, 10010187_6
KG131310, 10005659_26
GA0181577, 10001787_1
Kil181581, 10002095_17
Kil181584, 10003648_52
KG154502, 10003961_29
GJ24040145324, 10001267_3
KilMoana, 1000167_70
KilMoana, 1000372_10
KilMoana, 1000669_38
Mariner, 1000093_36
Mariner, 1000093_37
Mariner, 1000299_23
TahMoana, 1000070_25
TahMoana, 1002072_1
TubMallia, 1001167_2
GA0181567, 10001775_23
Kil181568, 10002680_23
Kil181569, 10001662_26
GA0181571, 10001624_23
GA0181571, 10001663_26

Candida, *maquilingensis*, J167
Candida, *Hydrothermarchoae*, archaea, JdFR_17JG2402035080, 001234
Candidatus, *Proteobacteria*, bacterium, RIFCSPHGW2, 01_16, 12
Candidatus, *Proteobacteria*, bacterium, RIFCSPLOW2, FULL_68, 19
COI, *202*, *FULL*, *Nitrospirae*, 41_54, curated
Chromola
Desulfibacter, *alkalitolans*
Moorella, *thermoaerica*, 2
Nitrospirae, bacterium, RIFCSPLOW2, 01_FULL_39, 10
Pyrobaculum, *arsenatum*, 1
Pyrobaculum, *arsenatum*, 2
Abe_10001651
Abe_10002020
Ga0052187_10001_184
Ga0052187_10002_1
Ga0052187_10007_1
Ga0052187_10045_82
Ga0052187_10045_83
Ga0052187_10048_80
Ga0052187_10048_81
Ga0052187_10081_115
Ga0052187_10083_108
Ga0052187_10086_102
Ga0052187_10089_127
Ga0052187_10090_49
Ga0052187_10091_49
Ga0052187_10092_52
Ga0098466_100117_1
Ga0098466_1001378_11
K0115000_10010187_6
Ga016311_10002569_26
Ga0181577_10001787_1
Ga0181581_10002035_17
Ga0181584_10000368_52
Ga0211543_10000631_39
JG2404014524_10012673
KiloMoana_10000167_70
KiloMoana_10000372_10
KiloMoana_10000689_38
Mariner_1000093_36
Mariner_1000093_37
Mariner_1000289_23
TarlMoana_1000070_25
TarlMoana_1000079_21
TurMoila_1001167_2
Ga0181567_10001775_23
Ga0181568_10002080_23
Ga0181569_10000166_26
Ga0181571_10001624_23
Ga0181572_10001693_1

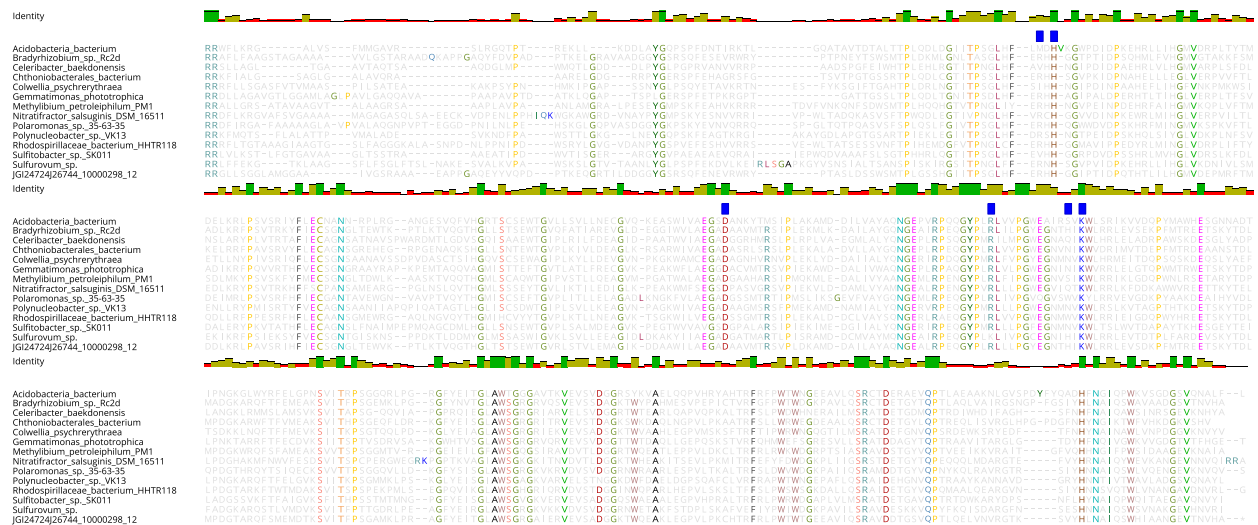
[illegible]

Supplementary Figure 1. DsrA Protein alignment and identified conserved residues in microbial and phage sequences. Highlighted amino acids indicate pairwise identity of $\geq 95\%$ and colored boxes indicate substrate binding motifs (pink) and strictly conserved siroheme binding motifs (blue). An identity graph (top) was fitted to the alignments to visualize pairwise identity at the following thresholds: 100% (green), 99-30% (yellow, scaled) and 29-0% (red, scaled).

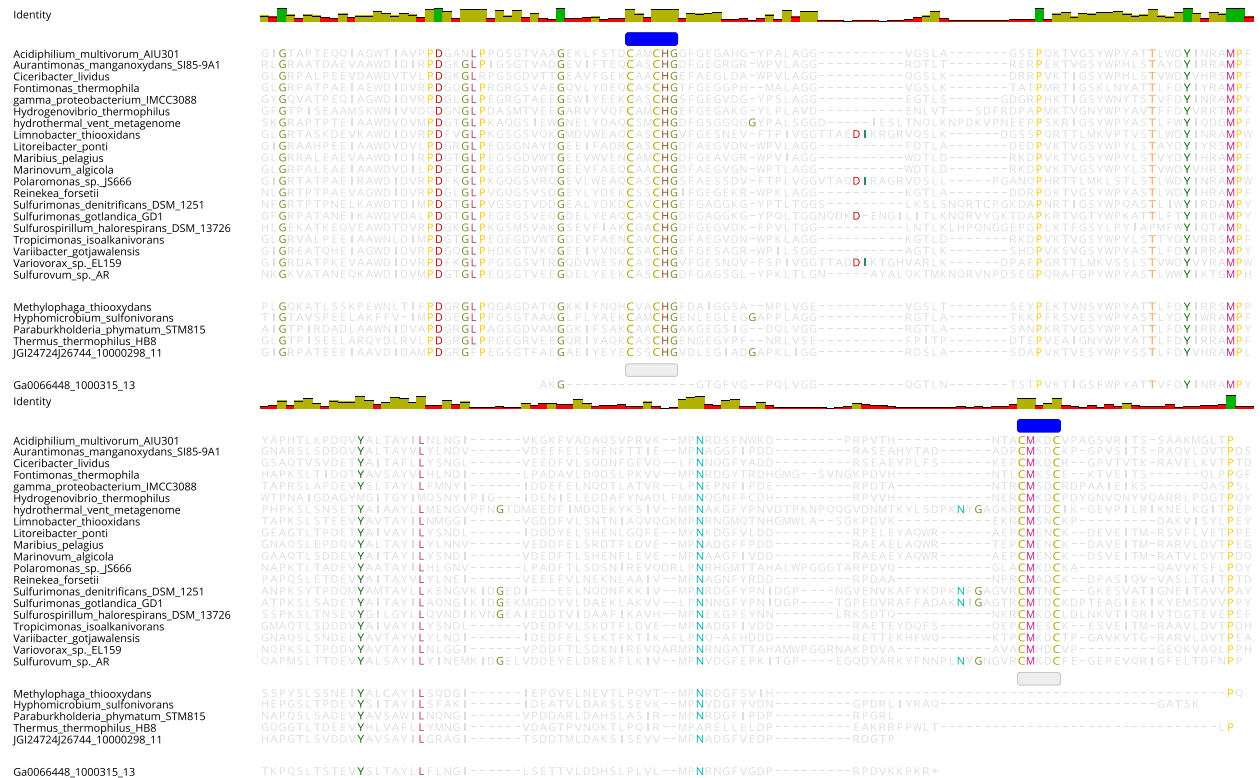




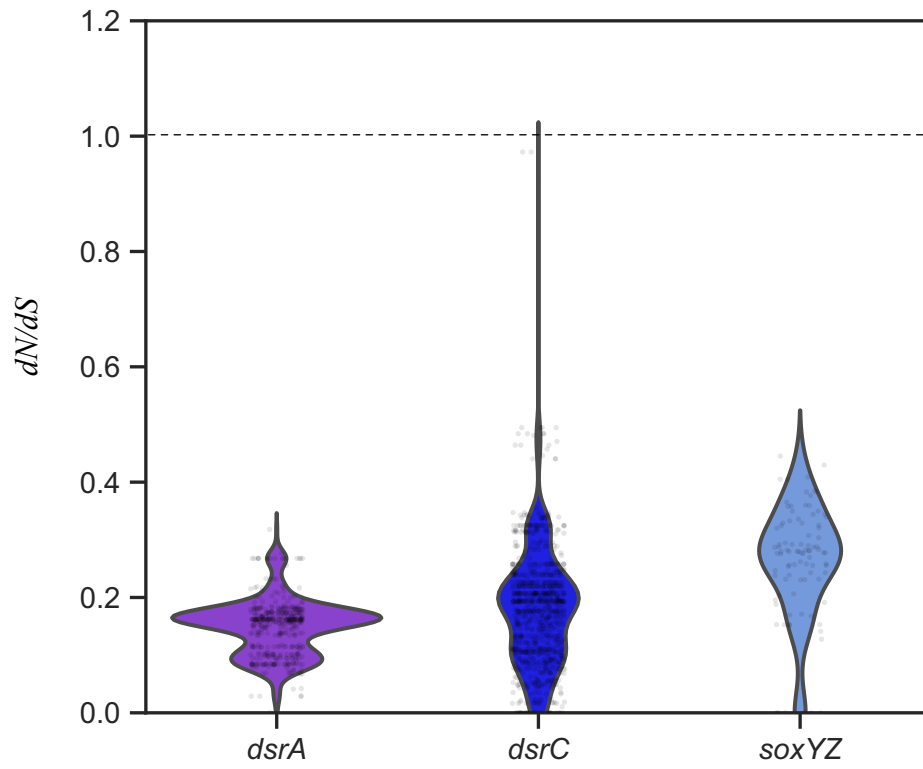
Supplementary Figure 3. SoxYZ protein alignment and conserved residues in microbial and phage sequences. Highlighted amino acids indicate pairwise identity of $\geq 90\%$ and colored boxes indicate substrate binding cysteine (blue) and cysteine motif (pink). An identity graph (top) was fitted to the alignments to visualize pairwise identity at the following thresholds: 100% (green), 99-30% (yellow, scaled) and 29-0% (red, scaled).



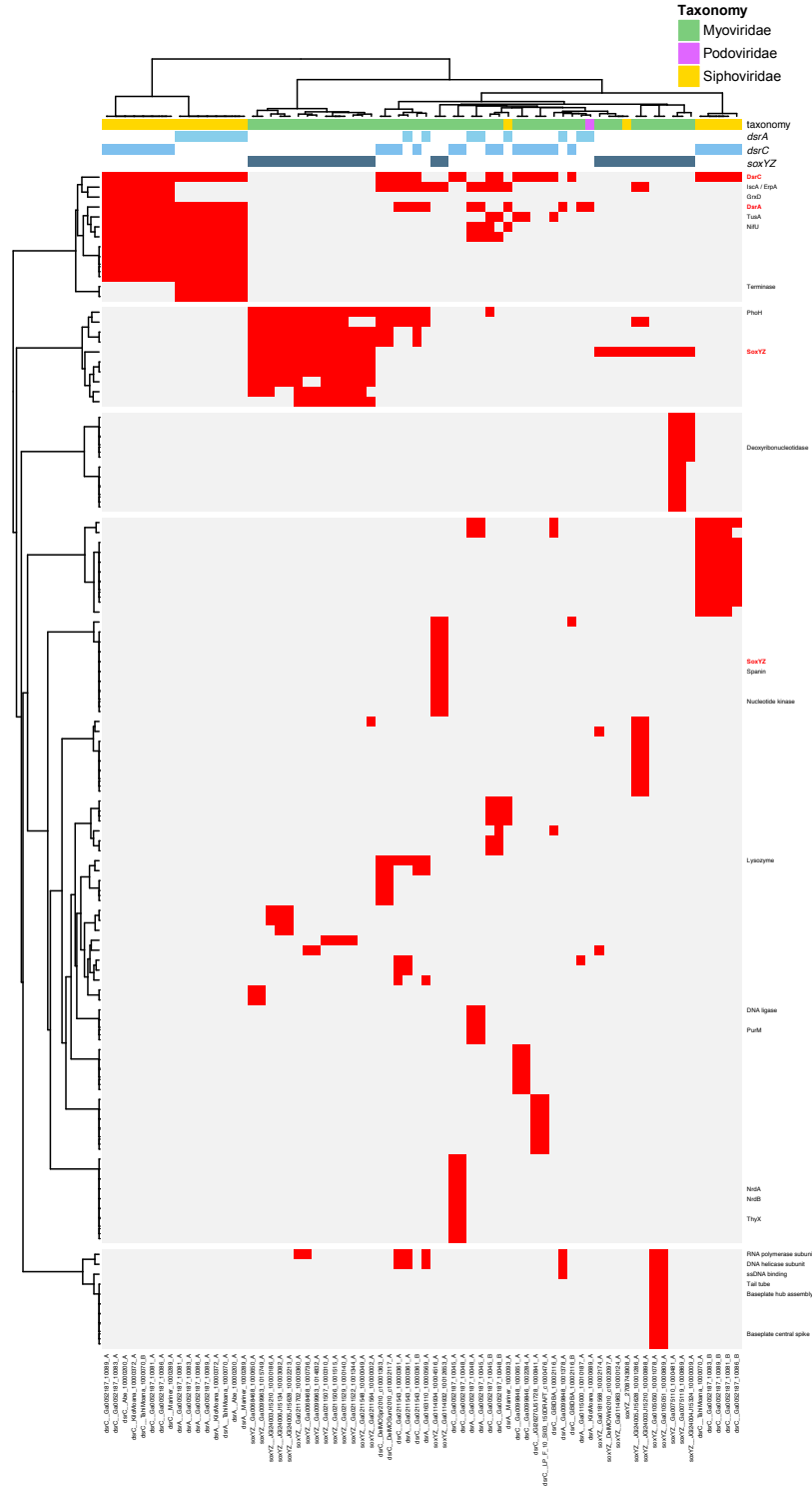
Supplementary Figure 4. SoxC protein alignment and conserved residues in microbial and phage sequences. Highlighted amino acids indicate pairwise identity of $\geq 90\%$ and colored boxes indicate cofactor coordination / active site (blue). An identity graph (top) was fitted to the alignments to visualize pairwise identity at the following thresholds: 100% (green), 99-30% (yellow, scaled) and 29-0% (red, scaled).



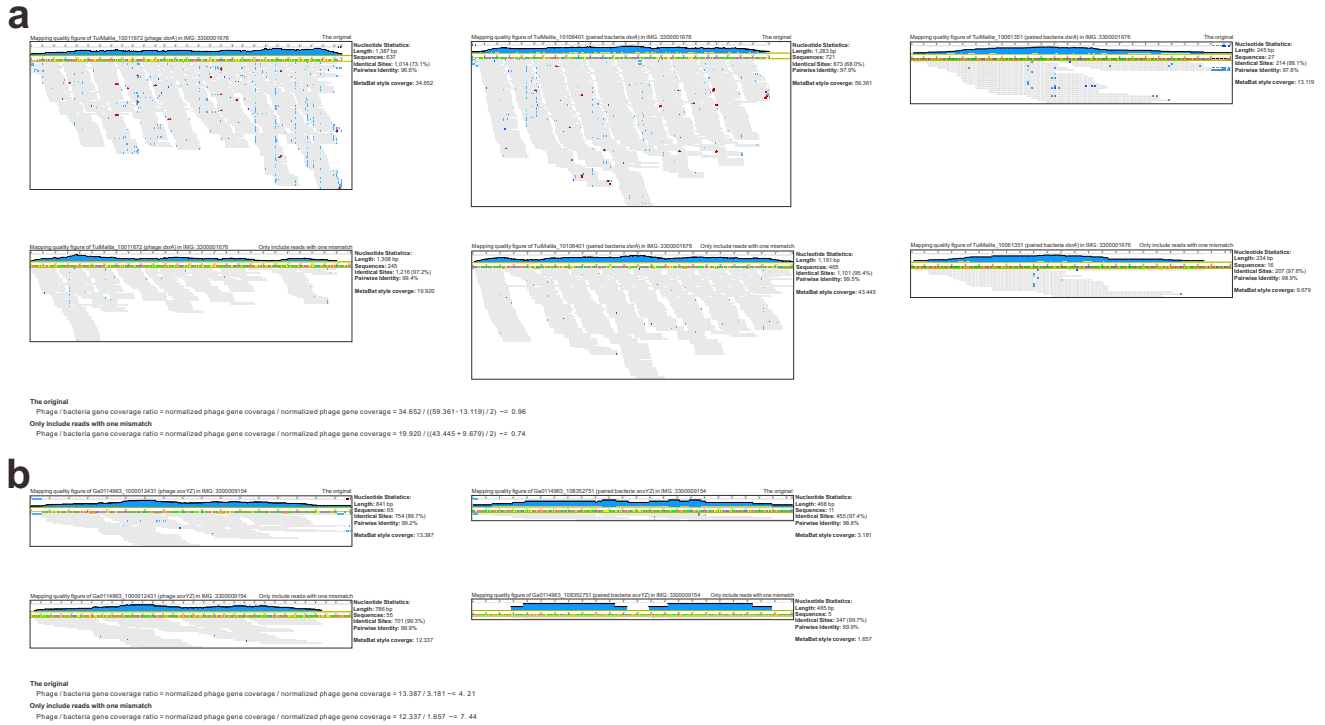
Supplementary Figure 5. SoxD protein alignment and conserved residues in microbial and phage sequences. Highlighted amino acids indicate pairwise identity of $\geq 95\%$ and colored boxes indicate cytochrome *c* motif (blue). An identity graph (top) was fitted to the alignments to visualize pairwise identity at the following thresholds: 100% (green), 99-30% (yellow, scaled) and 29-0% (red, scaled).



Supplementary Figure 6. Calculation of the ratio of non-synonymous to synonymous (dN/dS) nucleotide differences of AMGs. Comparison of dN/dS ratios between mVC AMG pairs for *dsrA*, *dsrC* and *soxYZ*. Each point represents a single comparison pair. Values below 1 suggest purifying selection pressures.



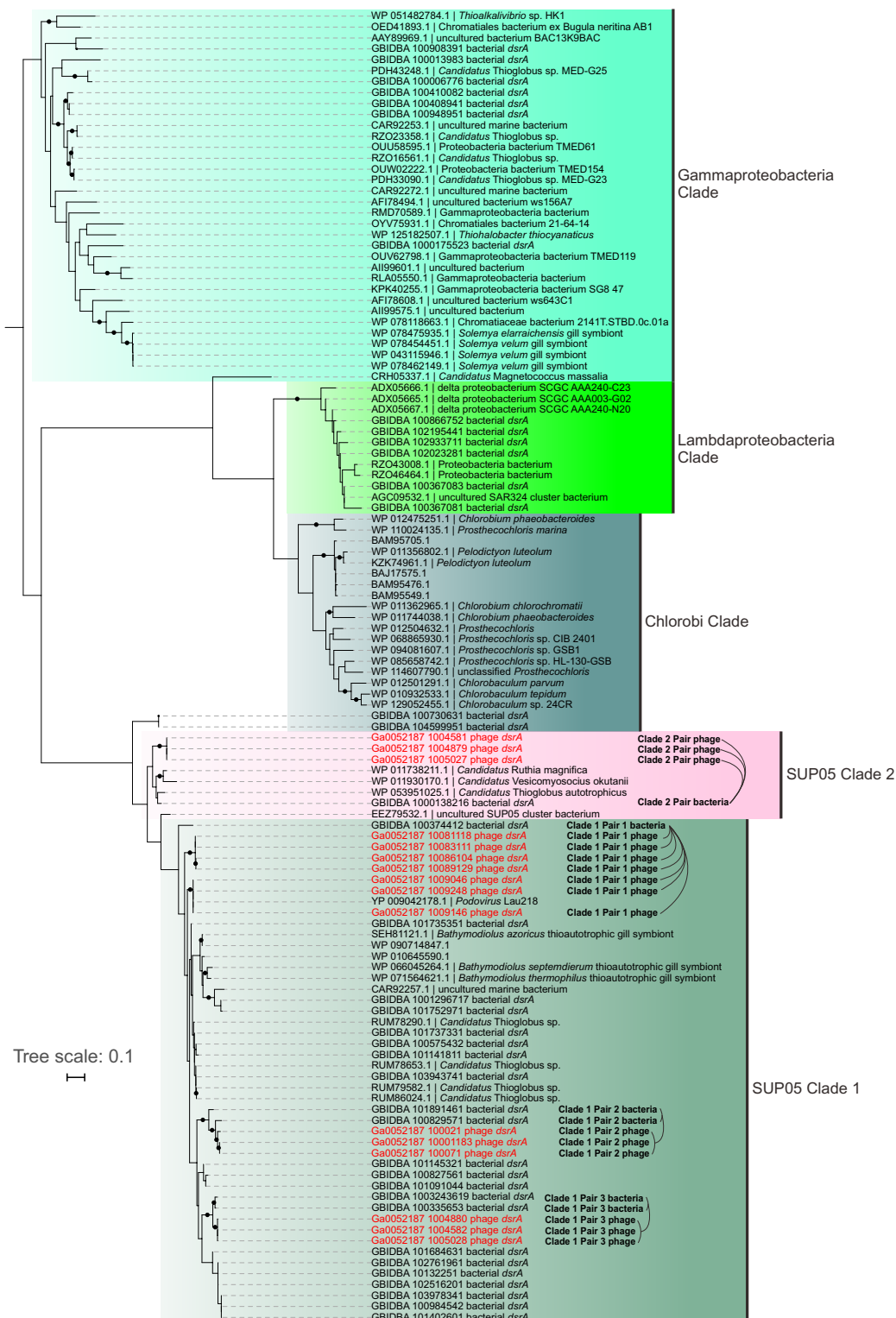
Supplementary Figure 7. Protein grouping of mVC AMG neighbor proteins. mVC hierarchical protein grouping where each row represents a single protein group (116 total) and each column represents a single mVC (70 total). For each mVC, 11 proteins were grouped, representing the encoded AMG plus five protein before and 5 proteins after the AMG. Metadata for encoded AMGs and estimated taxonomy per mVC is shown. mVC names are appended with “A” or “B” to distinguish those that encode multiple AMGs.



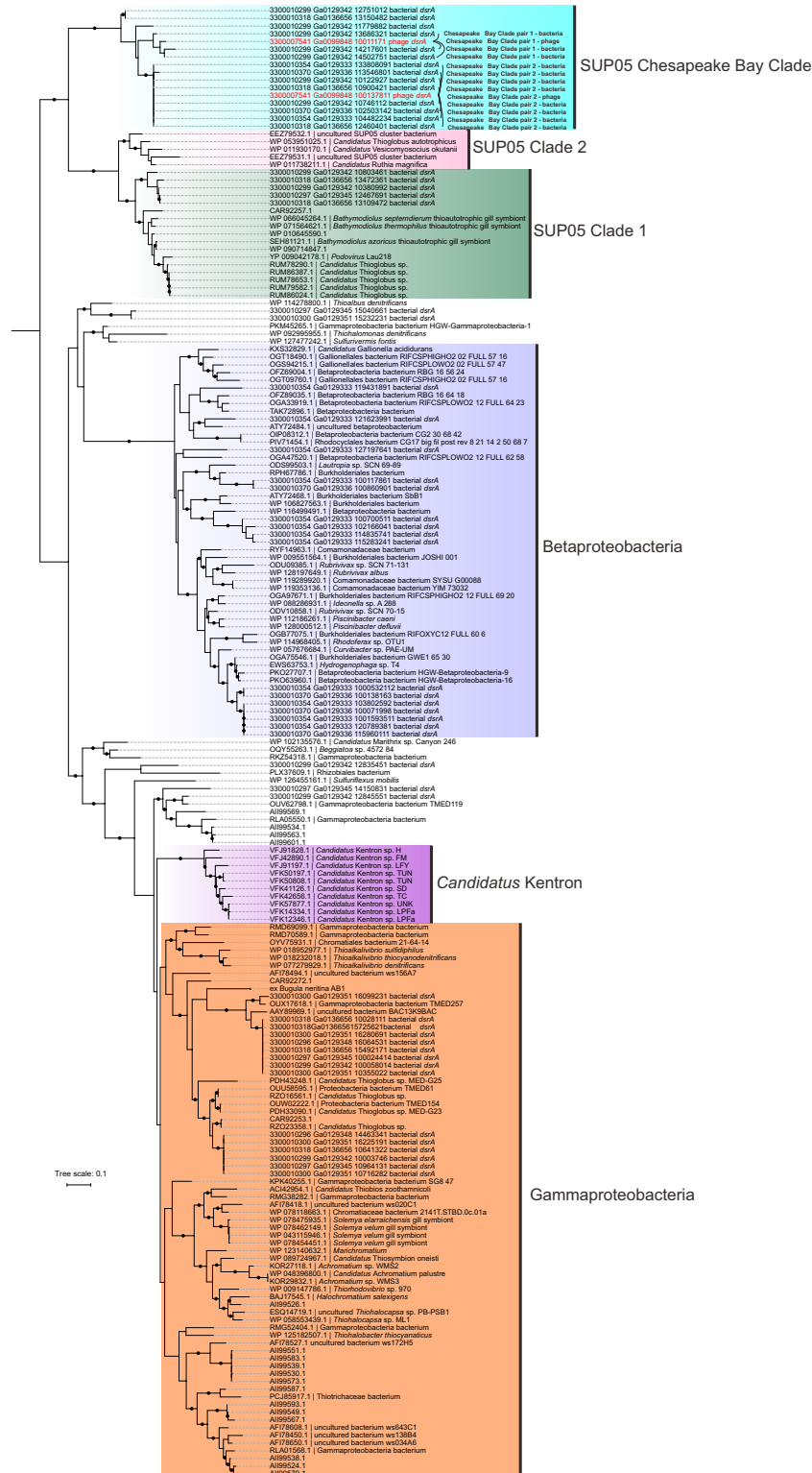
Supplementary Figure 8. Mapping quality checks for phage and bacterial sulfur AMGs. a Result for phage and bacterial *dsrA* genes in the metagenome IMG: 3300001676. The phage-host pair contains one phage *dsrA* (TuiMalila_10011672) and two bacterial *dsrA* (TuiMalila_10106401, TuiMalila_10061351). Both the original mapping result and the mapping results including reads with one mismatch were compared. The normalized phage / bacteria gene coverage ratios were calculated for both of the above settings. The normalized phage/bacteria gene coverage ratio based on the original mapping result are shown in Fig. 7a. **b** Result for phage and bacterial *soxYZ* gene in the metagenome of IMG: 3300009154. The phage-host pair contains one phage *soxYZ* (Ga0114963_1000012431) and one bacterial *soxYZ* (Ga0114963_108352751). Both the original mapping result and the mapping results including reads with one mismatch were compared. The normalized phage/bacteria gene coverage ratios were calculated for both of the above settings. The normalized phage/bacteria gene coverage ratios based on the original mapping results are shown in Fig. 7b. Filtering steps to only retain reads with only one mismatch was conducted by mapped.py (<https://github.com/christophertbrown/bioscripts/blob/master/ctbBio>) with the settings of "-m 1 -p both". Mapping results were visualized by Geneious Prime v2020.1.2.



Supplementary Figure 10. Phylogenetic tree of phage and bacterial SoxYZ and phage-host gene pairs from freshwater lake and *Tara* Ocean samples. The phage and bacterial *soxYZ* encoding proteins from the metagenomes studied in this project were aligned with reference sequences. The phylogenetic tree was reconstructed by IQ-TREE v1.6.9 with settings as described in the methods. Branches with over 90% UFBoot bootstrap values were labeled with closed circles. Phage *soxYZ* genes are labeled in red. The phage-host gene pairs (linked with dash lines) were labeled accordingly in the tree. The IMG metagenome IDs of freshwater lake and *Tara* Ocean samples are available in Supplementary Data 4.



Supplementary Figure 11. Phylogenetic tree of phage and bacterial DsrA and phage-host pairs from the Guaymas Basin hydrothermal environment. The phage and bacterial *dsrA* encoding proteins from the metagenomes studied in this project were aligned with reference sequences. The phylogenetic tree was reconstructed by IQ-TREE v1.6.9 with settings as described in the methods. Branches with over 90% UFBoot bootstrap values were labeled with closed circles. Phage *dsrA* genes are labeled in red. The phage-host gene pairs (linked with dash lines) were labeled accordingly in the tree. The IMG metagenome IDs of Guaymas Basin samples are 3300001683 and 3300003086.



Supplementary Figure 12. Phylogenetic tree of phage and bacterial *DsrA* and phage-host pairs from Chesapeake Bay. The phage and bacterial *dsrA* encoding proteins from the metagenomes studied in this project were aligned with reference sequences. The phylogenetic tree was reconstructed by IQ-TREE v1.6.9 with settings as described in the methods. Branches with over 90% UFBoot bootstrap values were labeled with closed circles. Phage *dsrA* genes are labeled in red. The phage-host gene pairs (linked with dash lines) were labeled accordingly in the tree. IMG metagenome IDs are: 3300010370, 3300010354, 3300010299, 3300010318, 3300010297, 3300010300, and 3300010296.

