Gulf of Mexico blue hole harbors high levels of novel microbial lineages
Supplemental Figures

Patin NV, Dietrich A, Stancil A, Quinan M, Beckler JS, Hall ER, Culter J, Smith CG, Taillefert M, Stewart FJ
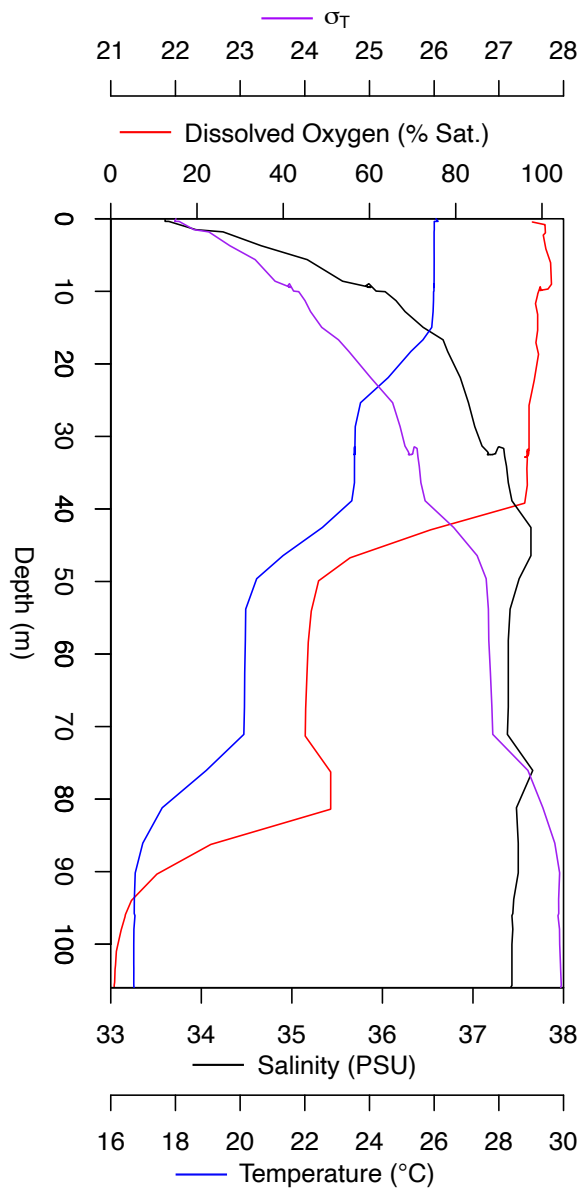
Supplemental Figures

Figure S1. Amberjack hole water column profile of salinity, temperature, density, and dissolved oxygen from May 2019.
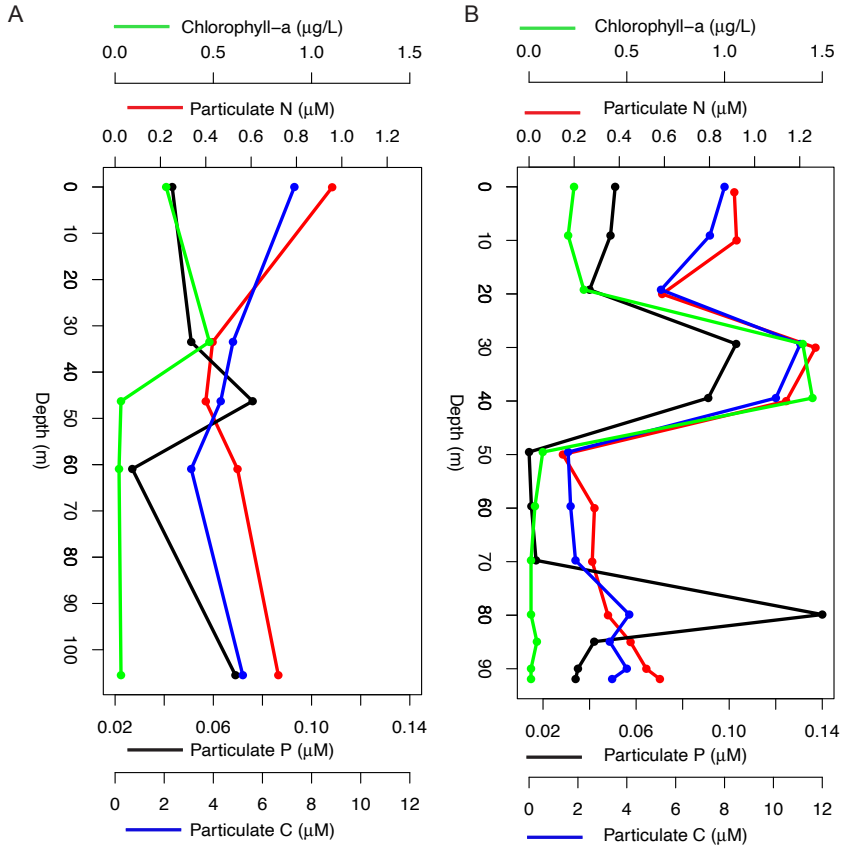
Figure S2. Chlorophyll and particulate nutrient profiles from May (A) and September (B) 2019. Points indicate the depths sampled for each measurement.
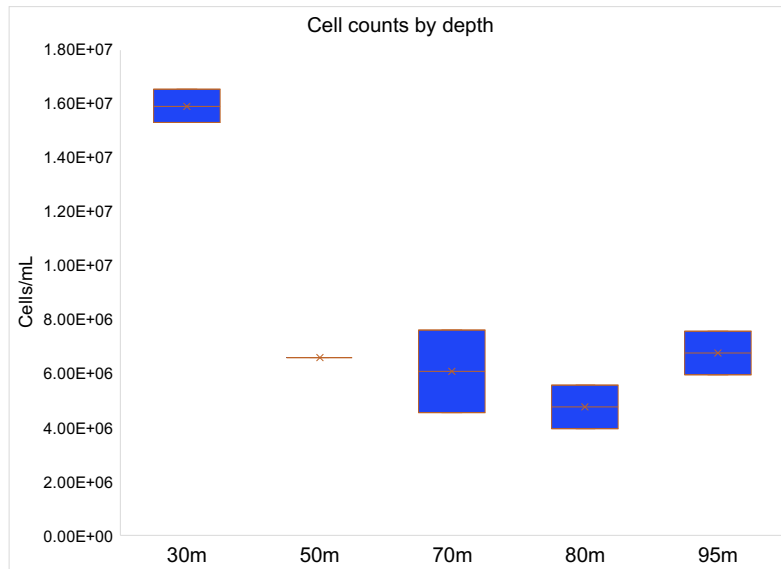

Figure S3. Average cell counts at five depth sampled in September 2019. All counts represent duplicate filters except 50 m which had only one filter.
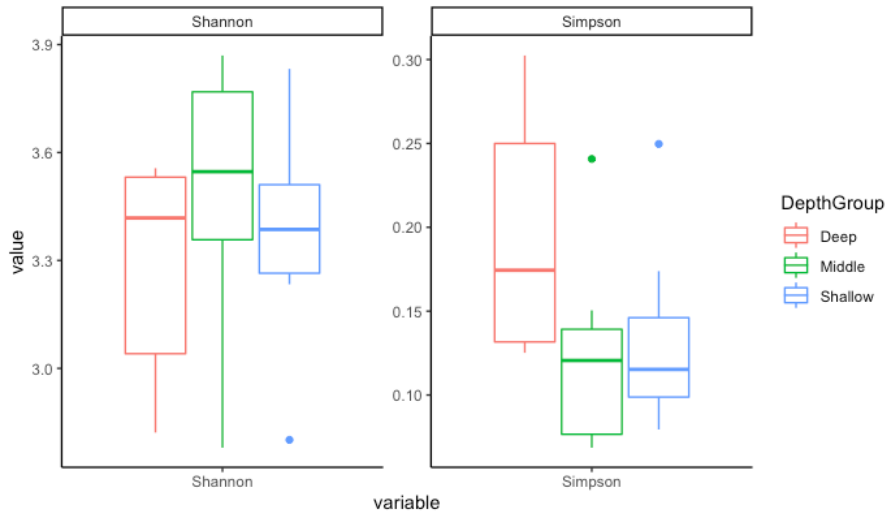
Figure S4. The medians and range of alpha diversity are comparable based on Shannon Diversity Index (*H*) across all depth groups; however, the Simpson Diversity Index (D) distinguishes the deep group from the other two depth groups.
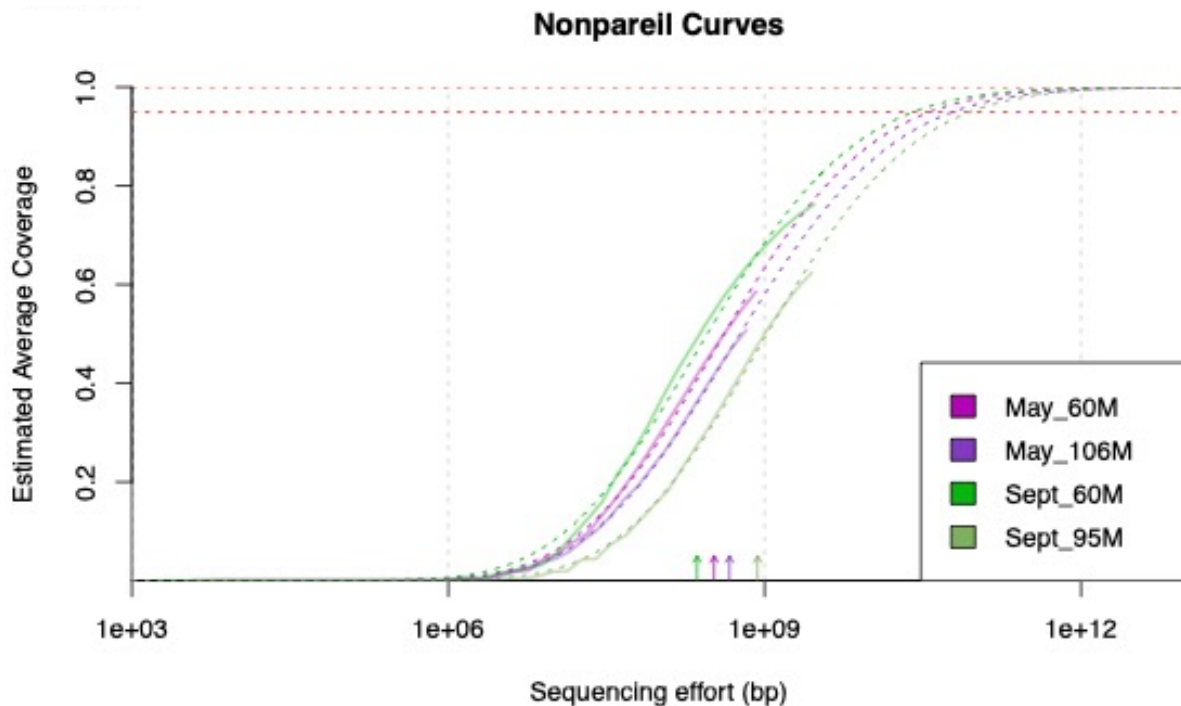

Figure S5. Nonpareil curves show estimated metagenome coverage as a function of the sequencing effort applied to each sample (May 60 m, May 106 m ("deep"), September 60 m, and September 95 m ("deep"). The arrows indicate the inflection point of the corresponding curvie, the log-natural of which is the Nonpareil Sequence Diversity Index (N_d). Lower Nd values indicate lower diversity, with less sequencing effort required to saturate the sequence space.
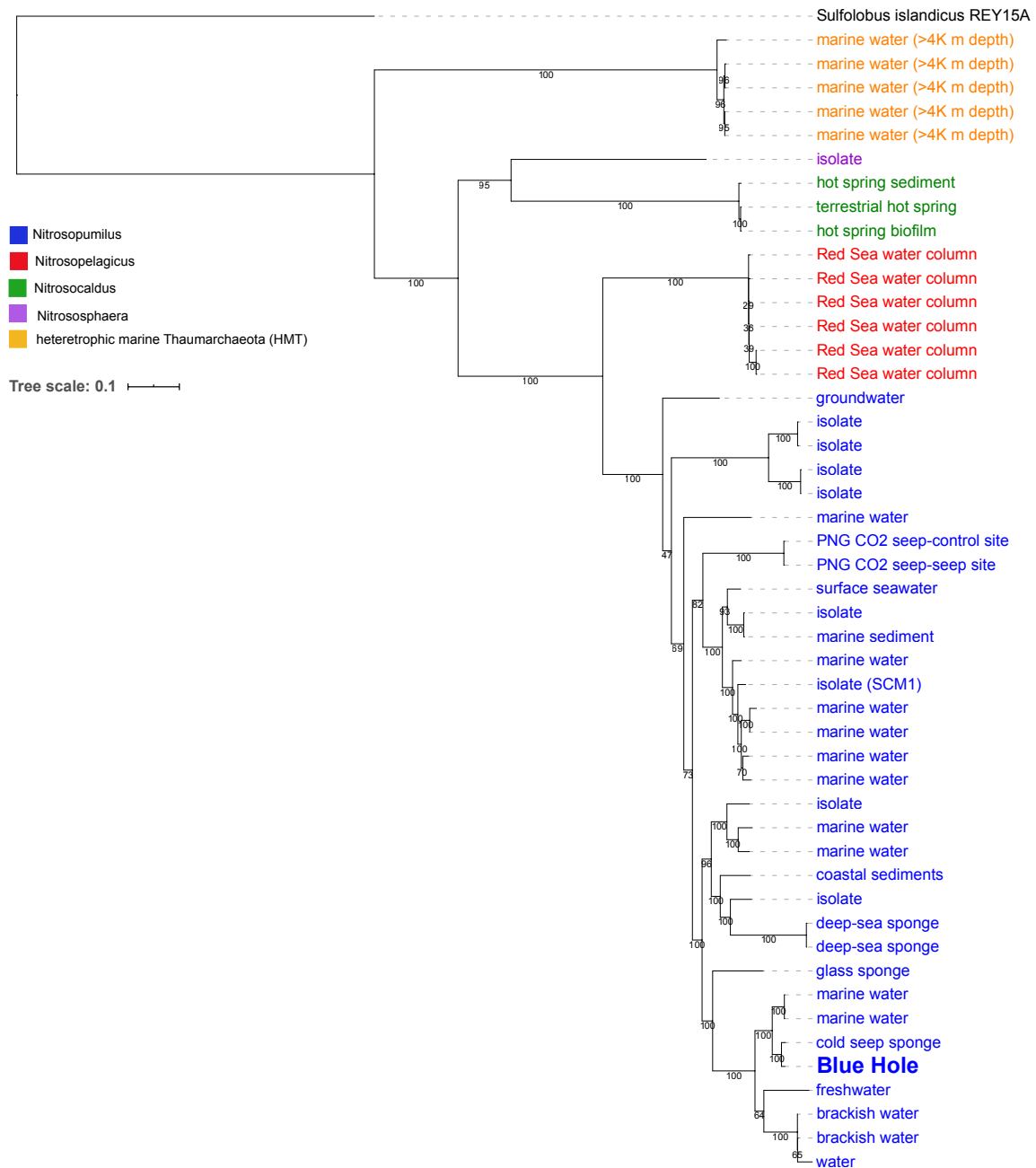
Figure S6. Maximum likelihood phylogenomic tree of the BH19 MAG places it among members of the genus *Nitrosopumilus*. Other Thaumarchaeotal clades shown include the genera *Nitrosopelagicus* (red), *Nitrosocaldus* (green), *Nitrososphaera* (purple), and the newly described heterotrophic marine Thaumarchaeota that do not perform ammonia oxidation, unlike all other representatives. Branch labels describe the strain or sample of origin. Bootstrap values (999 iterations) are shown for each node. BH19 is shown in bold lettering ("Blue Hole").
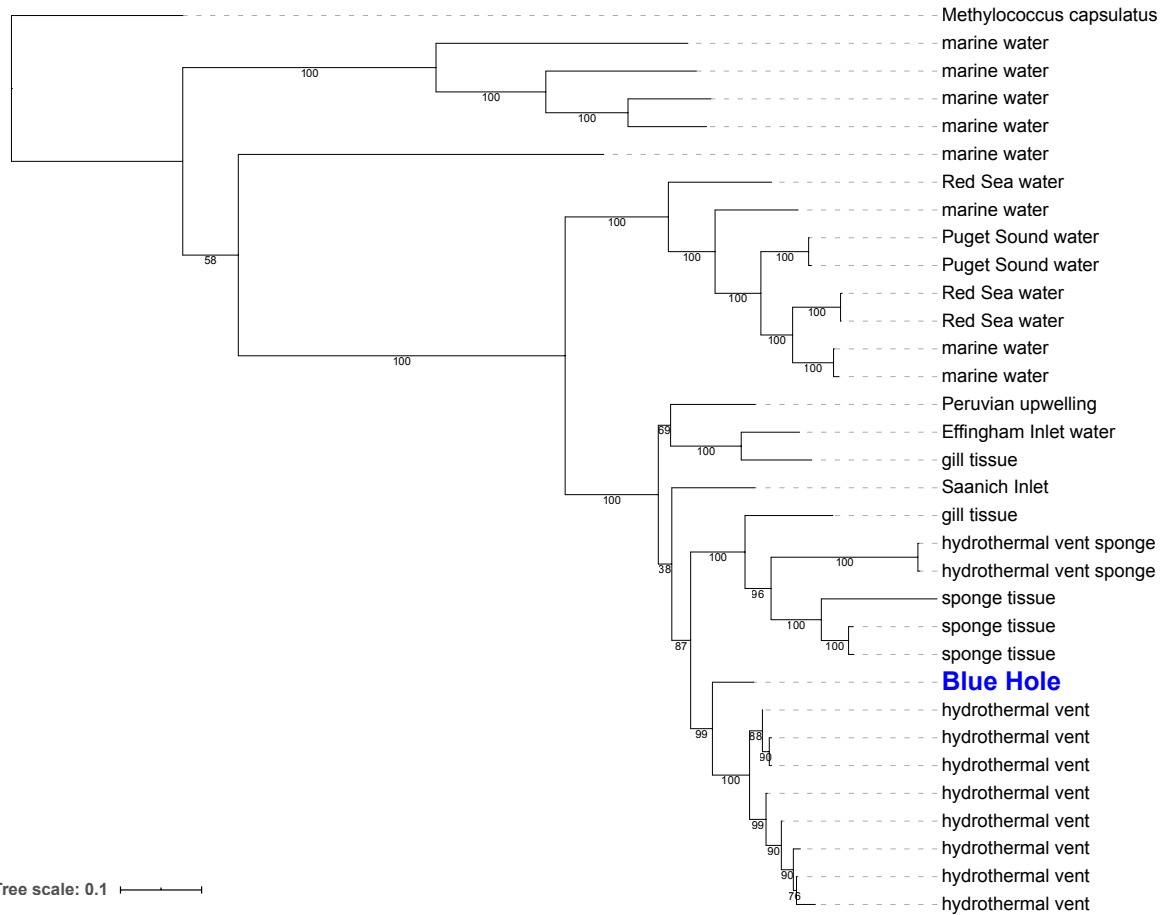
Figure S7. Maximum likelihood phylogenomic tree of the BH20 MAG places it in a sister branch to several *Thioglobus* sp. MAGs from hydrothermal vents. All reference genomes are members of the SUP05 clade (family *Thioglobaceae*). Branch labels describe the strain or sample of origin. Bootstrap values (999 iterations) are shown for each node. BH20 is shown in blue and bold lettering ("Blue Hole").

A. All KEGG 'Subgroups2' (222)
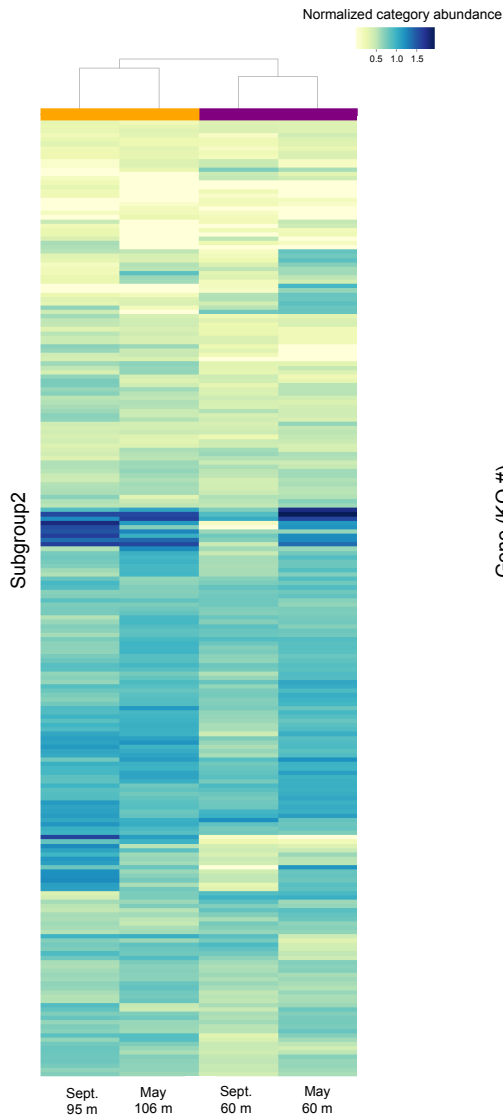
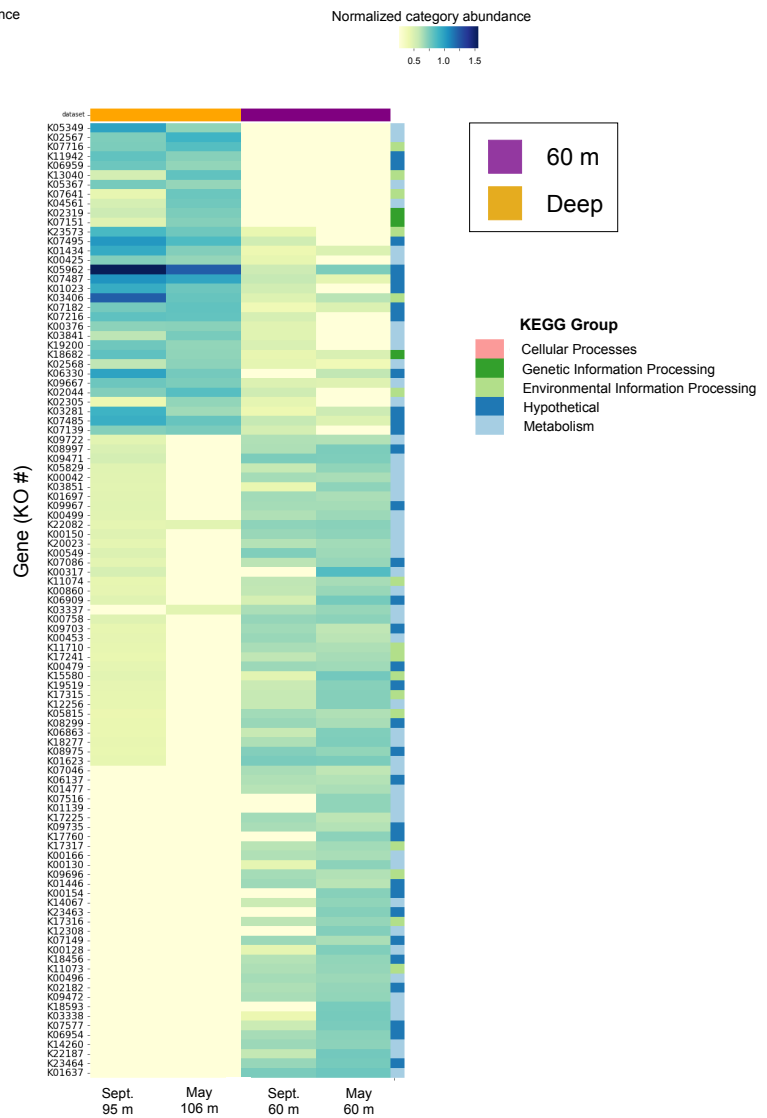**B. Genes with >0.2 average relative abundance and a difference between depth groups >20X**
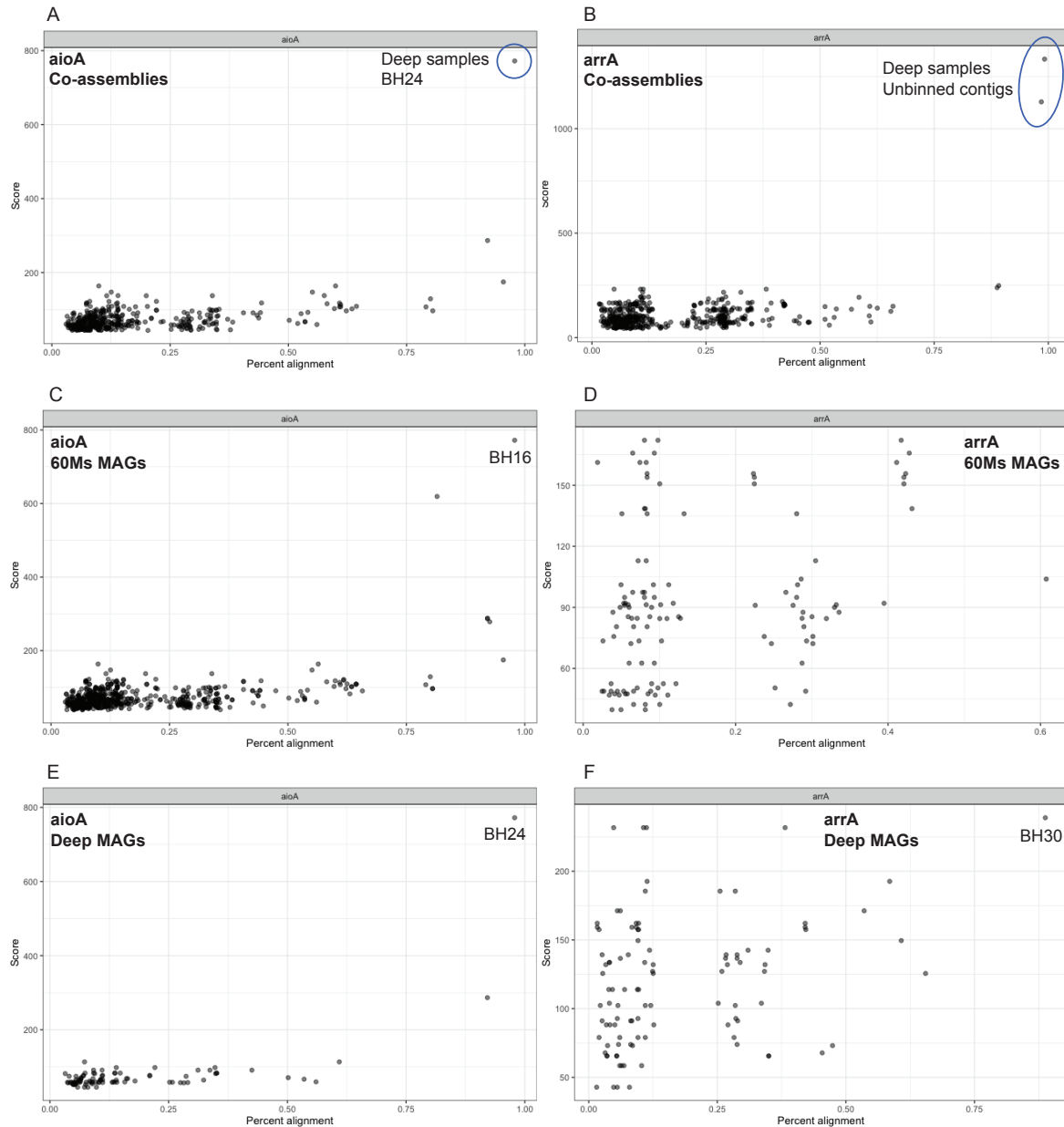

Figure S8. Heat maps show the relative abundance of KEGG categories and genes in the four individual metagenomic assemblies. (A) All categories at the second-highest hierarchical level ('Subgroup2,' one level above individual genes). (B) All genes with an average relative abundance greater than 0.2 and a difference between the deep and 60 m assemblies greater than 20X.

Figures S9. Quality assessments of the HMM results of genes queried for arsenic metabolism show presence of *aio*A (arsenite oxidase) and *arr*A (arsenate reductase) genes in assemblies and MAGs. (A, B) Open reading frames from both co-assemblies showed high-quality hits for each gene, with *aio*A located on a contig from the deep samples which was laso part of the BH16 MAG and *arr*A found only on unbinned contigs. (C, D) *aio*A was found in BH16, while there were no high-quality hits for *arr*A in the 60 m MAGs. (E, F) Queries against the deep MAGs showed aioA BH24 and *arr*A in BH30. The latter was confirmed by running the open reading frame in question against a functional gene database using blastp (Dunivin et al. 2019).

Supplemental Table Legends

Table S1. Amberjack Hole water column sample metadata and amplicon sequencing results. Processed reads are those that were quality-filtered, trimmed, and run through DADA2.

Table S2. Metagenome assembly data including sample source, pre- and post-quality filtered reads, and assembly statistics for both single and co-assemblies.

Table S3. Metagenome-assembled genomes (MAGs) and their associated samples, quality metrics, and assigned taxonomy.

Table S4. Metagenome-assembled genomes (MAGs) assessed in the phylogenomic trees (Fig. 5, Fig. S6, Fig. S7) including the Amberjack MAG of interest and the source of all other MAGs included in the phylogeny.