

Supplementary Material

1 SUPPLEMENTARY DATA

- **Supplementary Material 1.** Gene and pathways associated with each osmolyte included in this study. For each compound, the type of reaction, named reaction type, reaction number, enzyme number, and KEGG ortholog are detailed, as well as the the pathway and step in each pathway that were manually annotated for this study.
- **Supplementary Material 2.** KEGG ortholog presence and absence for each of the reference genomes from MarRef and transcriptomes within the MMETSP.
- **Supplementary Material 3.** Predicted synthesis, breakdown, and transport of each metabolite for each of the reference genomes from MarRef and transcriptomes from MMETSP.
- **Supplementary Material 4.** The HMM bitscore thresholds for each KEGG ortholog used in the analysis of recovered Tara orthologs from OM-RGC v2 and MATOU v1. The number of orthologs returned from Ocean Gene Atlas are reported (with an e-value cutoff of 1e-10). The orthologs that pass the HMM bitscore threshold were included in the presented analysis.

2 SUPPLEMENTARY FIGURES

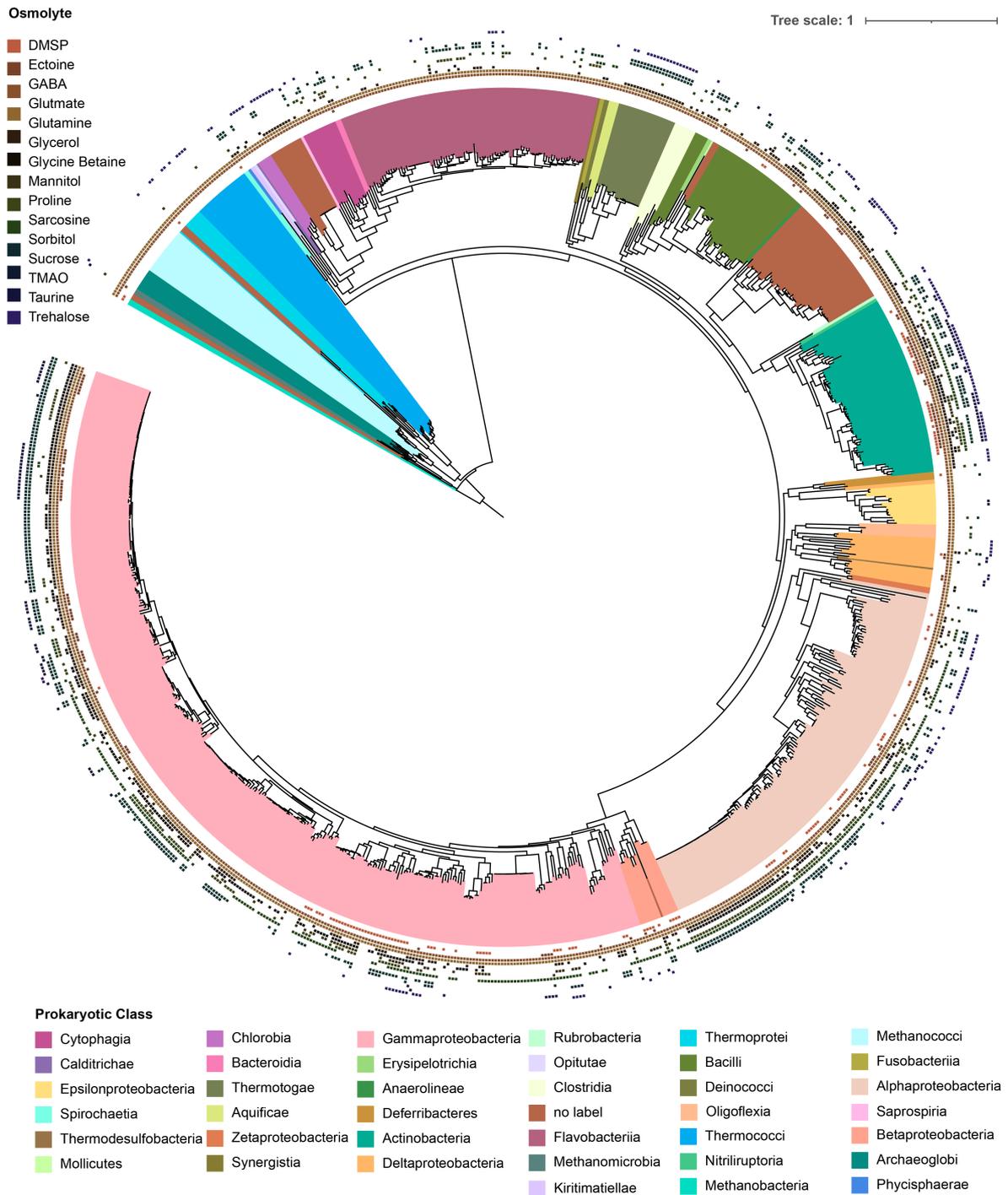


Figure S1. Maximum likelihood phylogeny of the MarRef genomes depicting synthesis of osmolytes of interest. The concatenated gene tree was constructed using the ‘Universal’ dataset of 15 genes as defined by Hug et al. (2016) and default parameters within GToTree. The presence or absence of the ability to synthesis of each of the osmolytes is depicted for each genome.

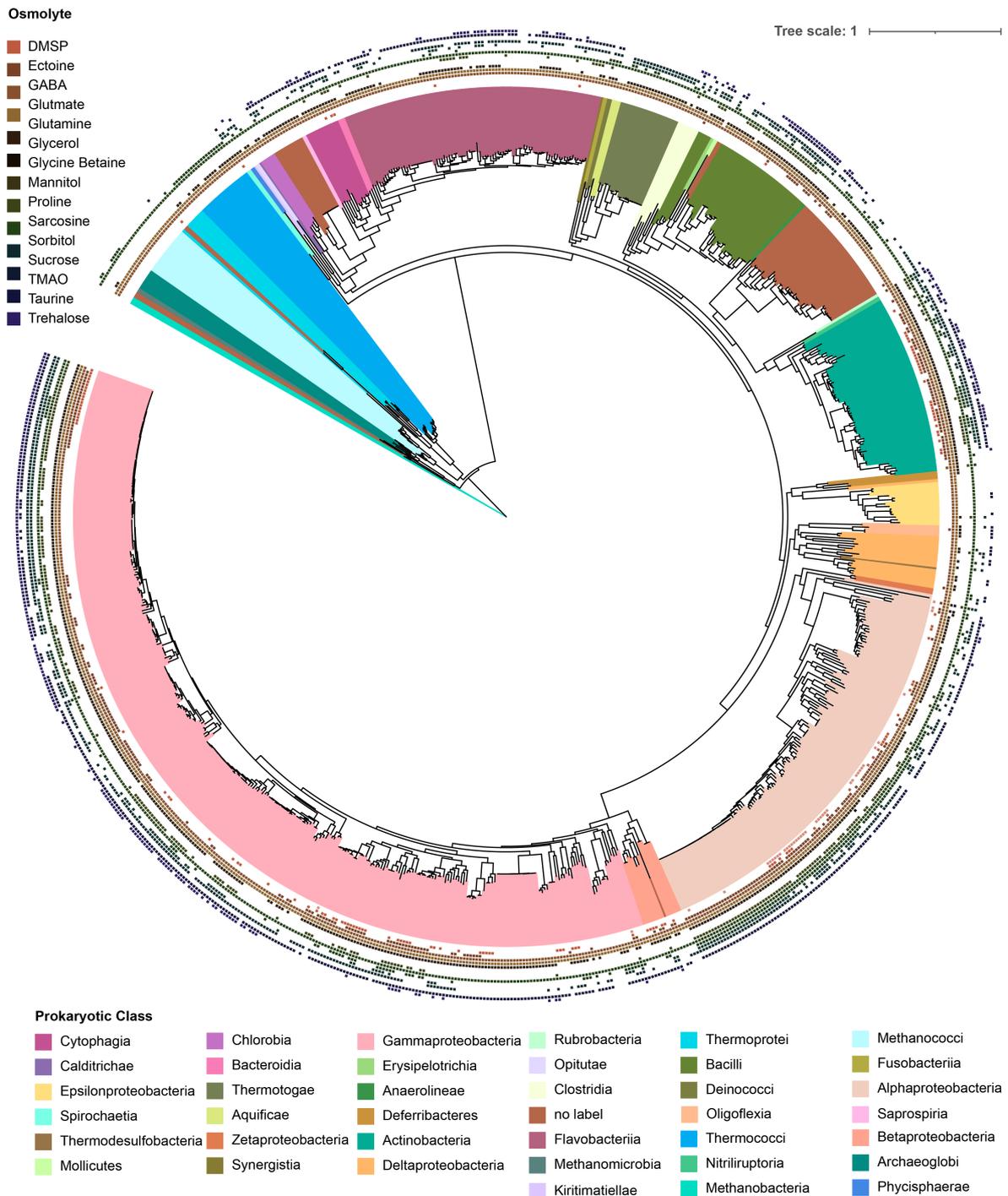


Figure S2. Maximum likelihood phylogeny of the MarRef genomes depicting breakdown of osmolytes of interest. The concatenated gene tree was constructed using the ‘Universal’ dataset of 15 genes as defined by Hug et al. (2016) and default parameters within GToTree. The presence or absence of the ability to breakdown of each of the osmolytes is depicted for each genome.

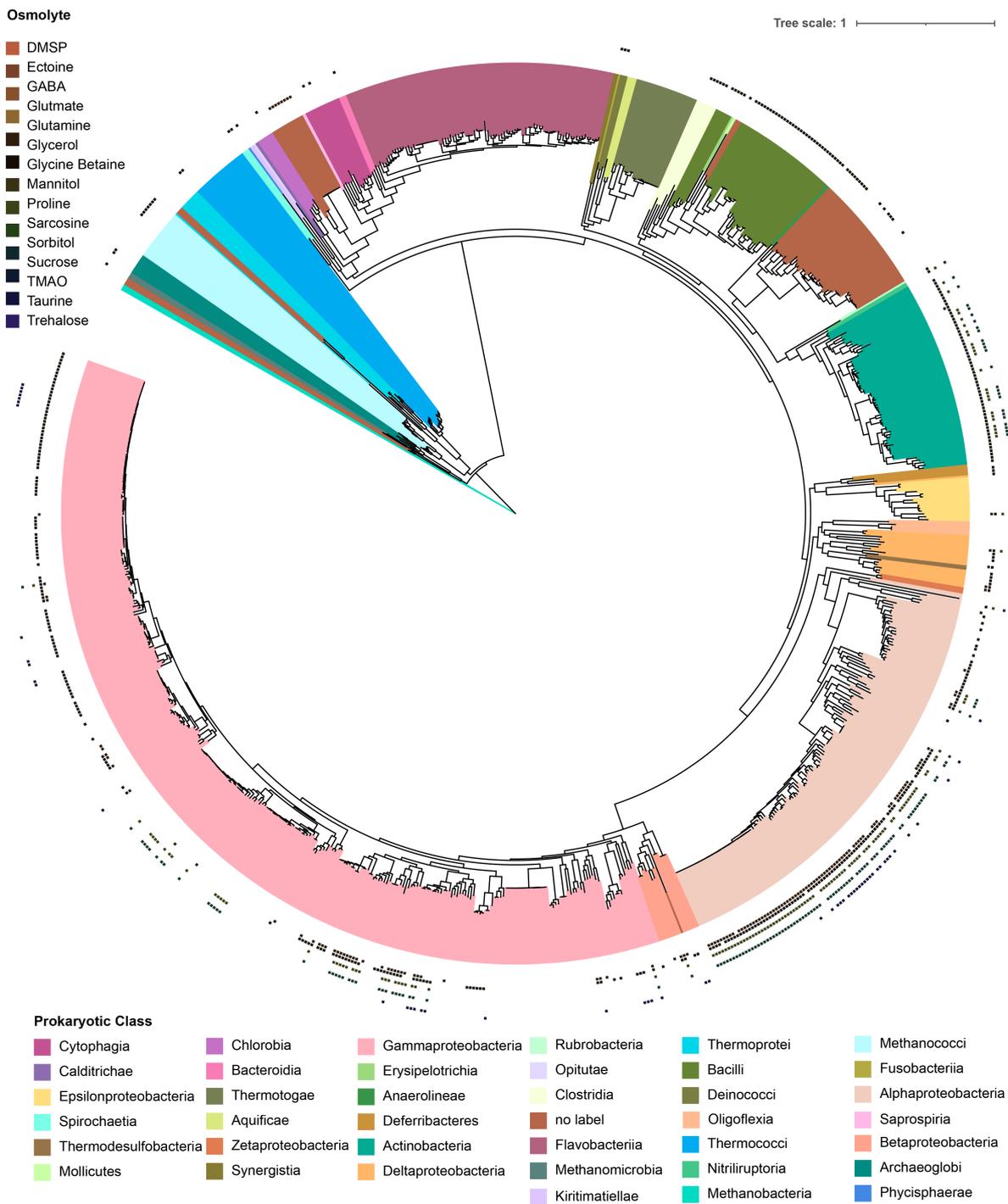


Figure S3. Maximum likelihood phylogeny of the MarRef genomes depicting transport of osmolytes of interest. The concatenated gene tree was constructed using the ‘Universal’ dataset of 15 genes as defined by Hug et al. (2016) and default parameters within GToTree. The presence or absence of the ability to transport of each of the osmolytes is depicted for each genome.

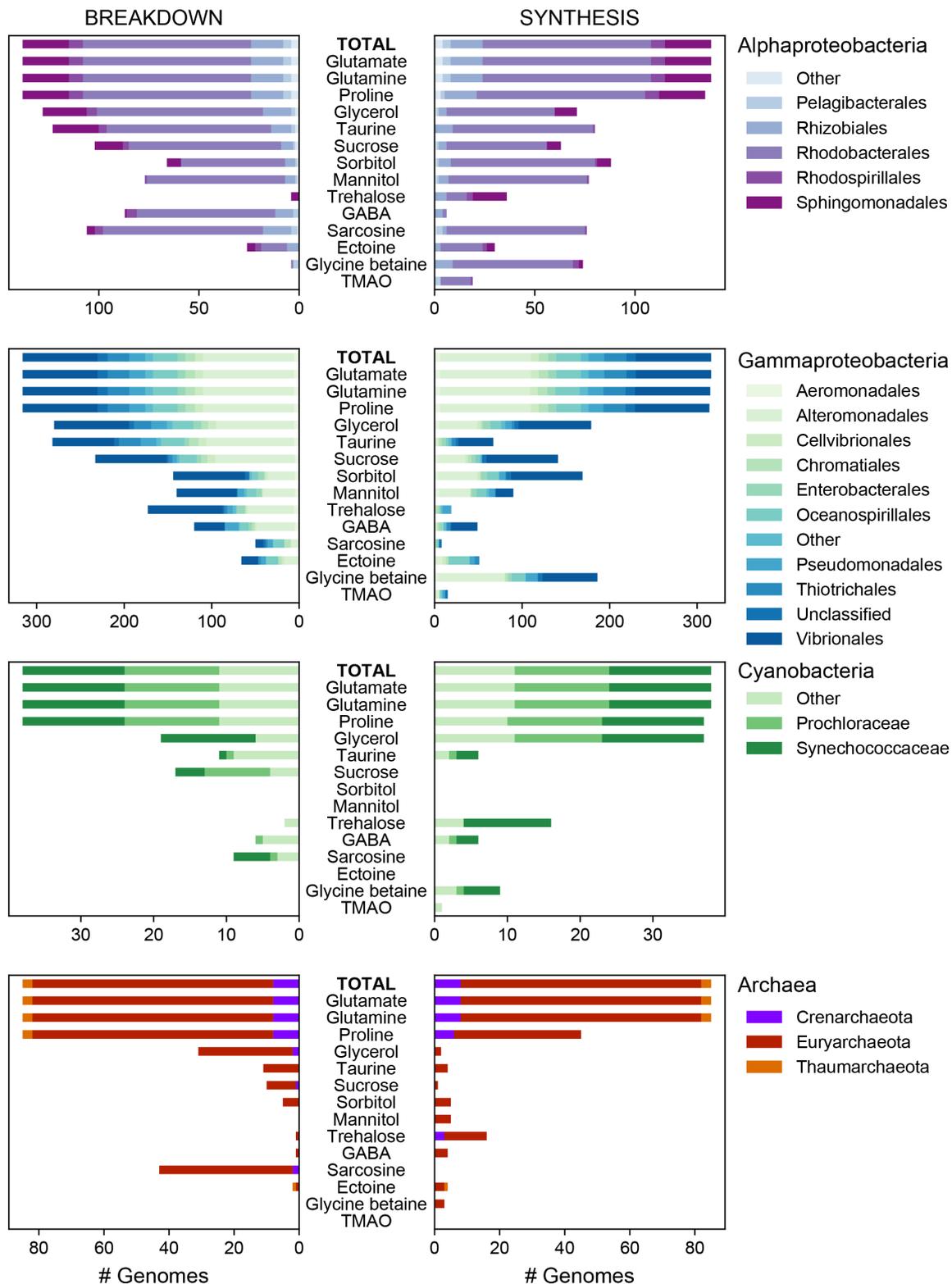


Figure S4. The predicted synthesis and breakdown of targeted osmolytes four major subgroups of prokaryotes shown in (Figure 1): Alphaproteobacteria, Gammaproteobacteria, Cyanobacteria, and Archaea. The breakdown and synthesis of osmolytes within these groups is depicted as a stacked bar graph, and is colored by the designated taxonomic order for Alpha- and Gammaproteobacteria, family for Cyanobacteria, and phylum for Archaea. Osmolytes are sorted along the y-axis based on the total number of genomes capable of breaking down a given osmolyte as in (Figure 1).

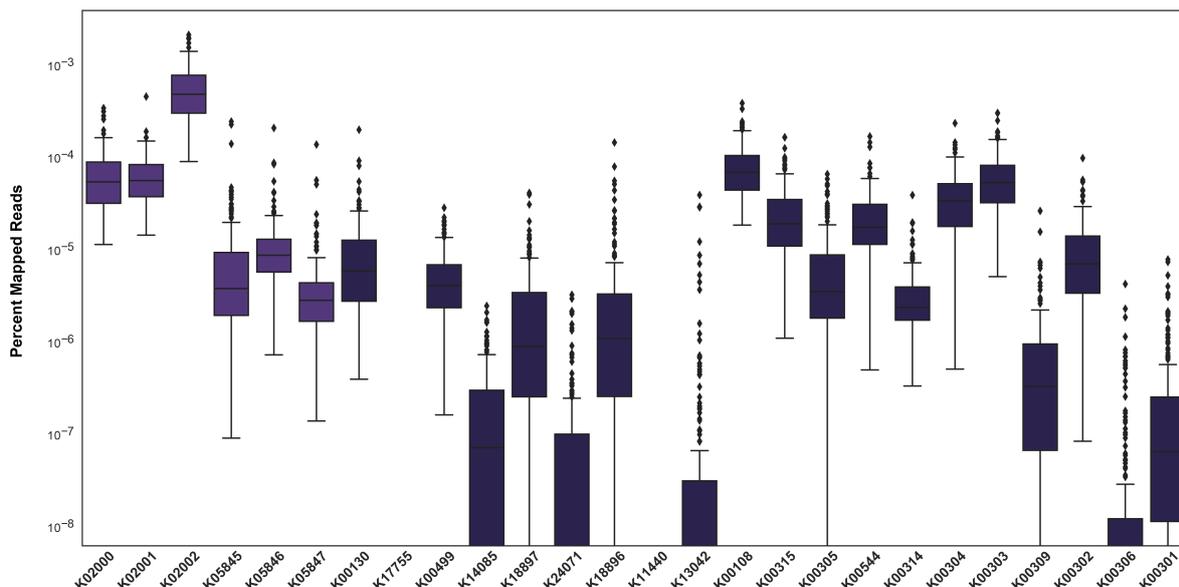


Figure S5. Prokaryotic metatranscriptome abundance of KEGG orthologs involved in the metabolism of glycine betaine. The percent mapped read abundance of all orthologs recovered for each KO were summed by sample for the bacterial metatranscriptomic data from Tara Oceans using the OM-RGC v2 dataset. The distribution of all abundance data is shown as a box plot. Transporters are highlighted in light purple, synthesis and breakdown KOs are highlighted in dark purple.

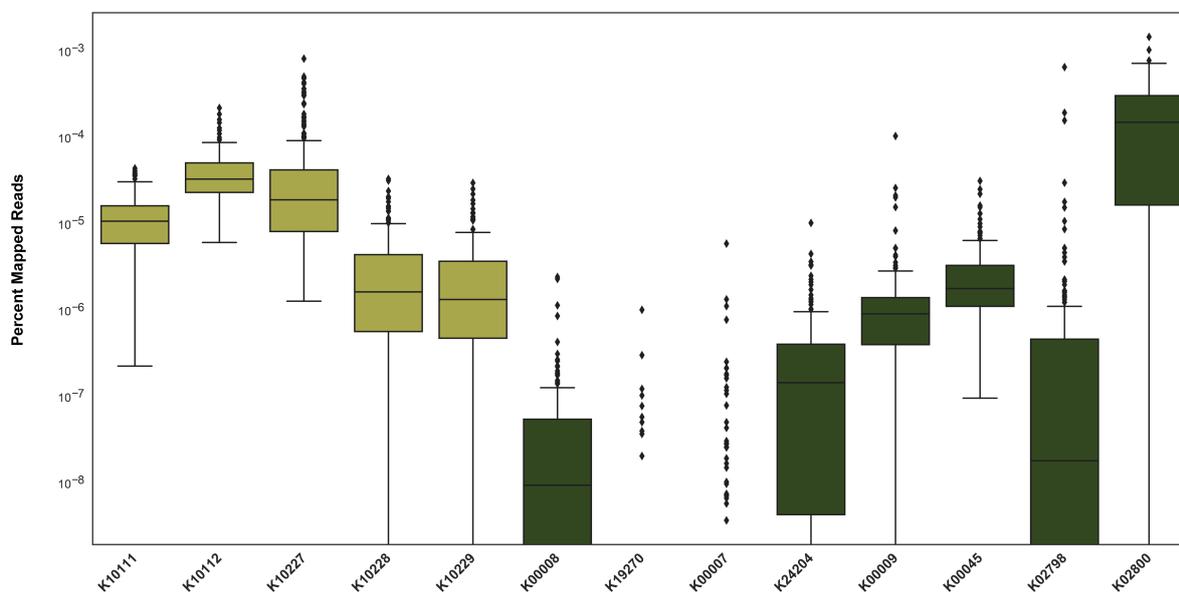


Figure S6. Prokaryotic metatranscriptome abundance of KEGG orthologs involved in the metabolism of mannitol. The percent mapped read abundance of all orthologs recovered for each KO were summed by sample for the bacterial metatranscriptomic data from Tara Oceans using the OM-RGC_v2 dataset. The distribution of all abundance data is shown as a box plot. Transporters are highlighted in light green, synthesis and breakdown KOs are highlighted in dark green.

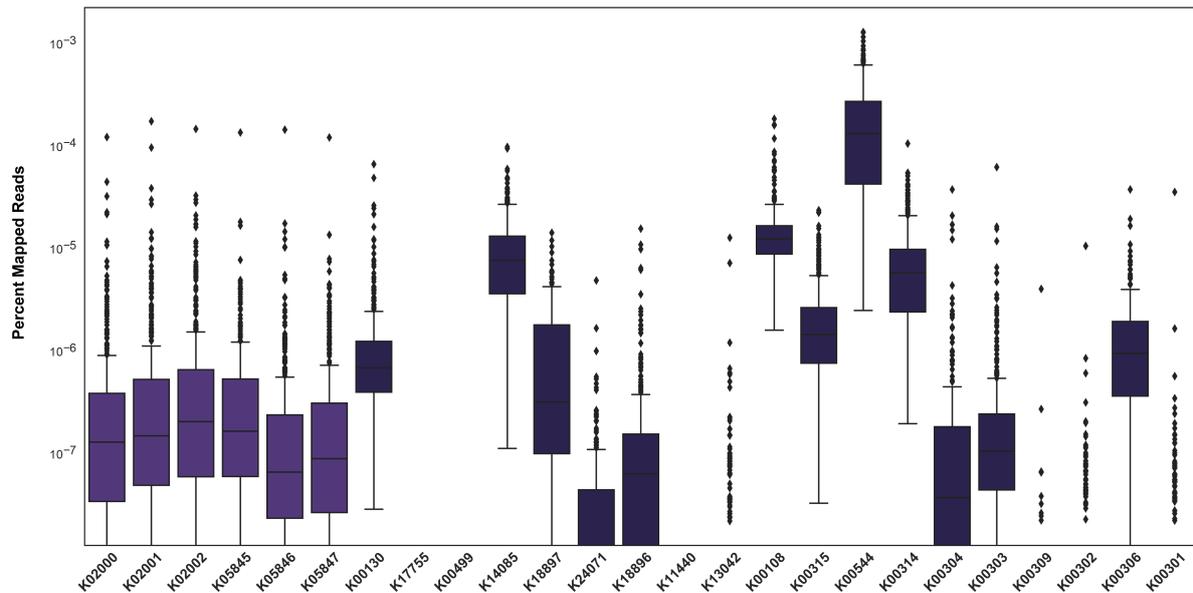


Figure S7. Eukaryotic metatranscriptome abundance of KEGG orthologs involved in the metabolism of glycine betaine. The percent mapped read abundance of all orthologs recovered for each KO were summed by sample for the bacterial metatranscriptomic data from Tara Oceans using the MATOU-v1 dataset. The distribution of all abundance data is shown as a box plot. Transporters are highlighted in light purple, synthesis and breakdown KOs are highlighted in dark purple.

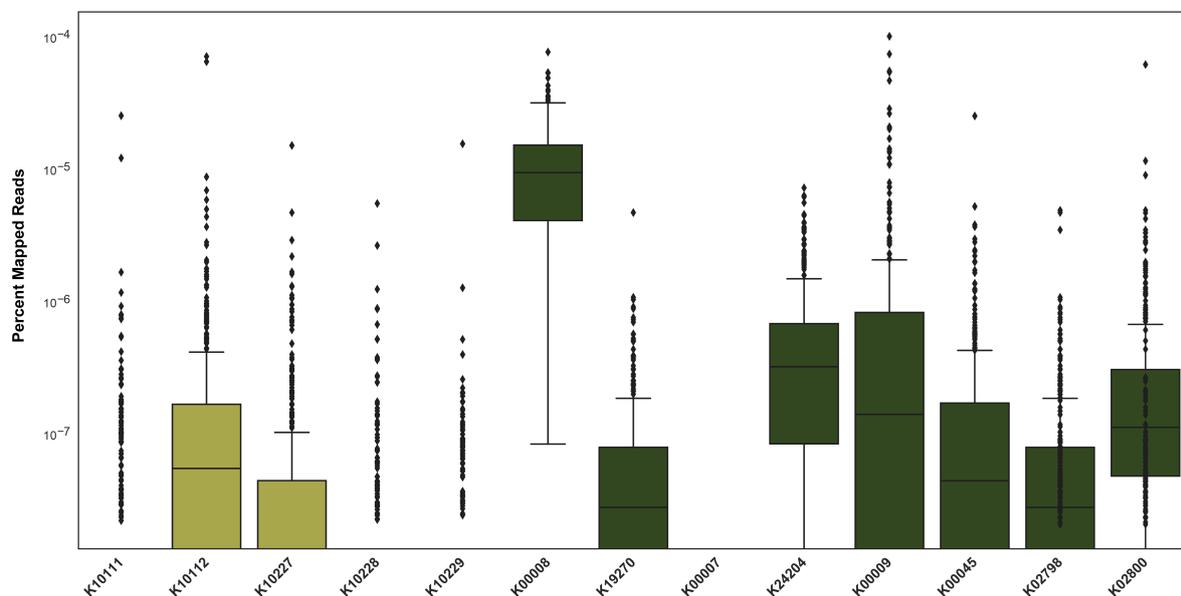


Figure S8. Eukaryotic metatranscriptome abundance of KEGG orthologs involved in the metabolism of mannitol. The percent mapped read abundance of all orthologs recovered for each KO were summed by sample for the bacterial metatranscriptomic data from Tara Oceans using the MATOU-v1 dataset. The distribution of all abundance data is shown as a box plot. Transporters are highlighted in light green, synthesis and breakdown KOs are highlighted in dark green.

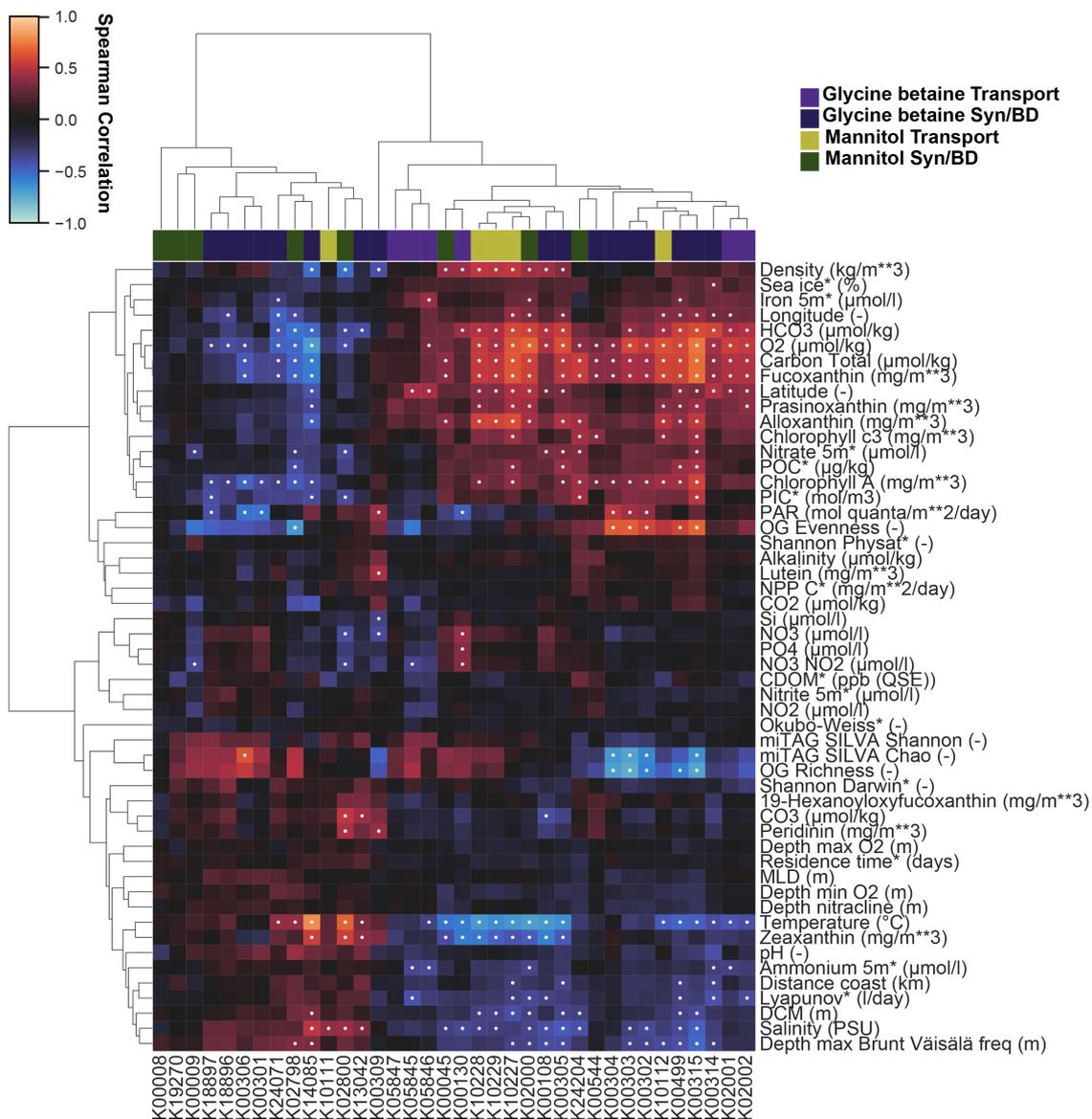


Figure S9. Spearman correlation against environmental parameters from Tara. The global abundance of key orthologs involved in the synthesis, breakdown, and transport of mannitol (green) and glycine betaine (purple) from prokaryotic metatranscriptomic data from Tara was correlated with core environmental parameters. The metatranscriptomic abundance (percent mapped reads) of each of the orthologs involved in the processing of the osmolytes shown here was assessed across the bacterial metatranscriptomic data from Tara Oceans using the OM-RGC_v2 dataset. The relative correlation of metatranscriptomic profiles for each of the orthologs considered was assessed by taking the Spearman’s correlation of the sum of all orthologs associated with a KO at a given site and correlating it with available metadata. The Spearman’s ρ is depicted with a heatmap that is clustered based on Bray-Curtis similarity. Significant correlations were determined with a Bonferroni multi-testing p-value correction, and significant correlations ($p < 0.001$) are depicted as a white dot.

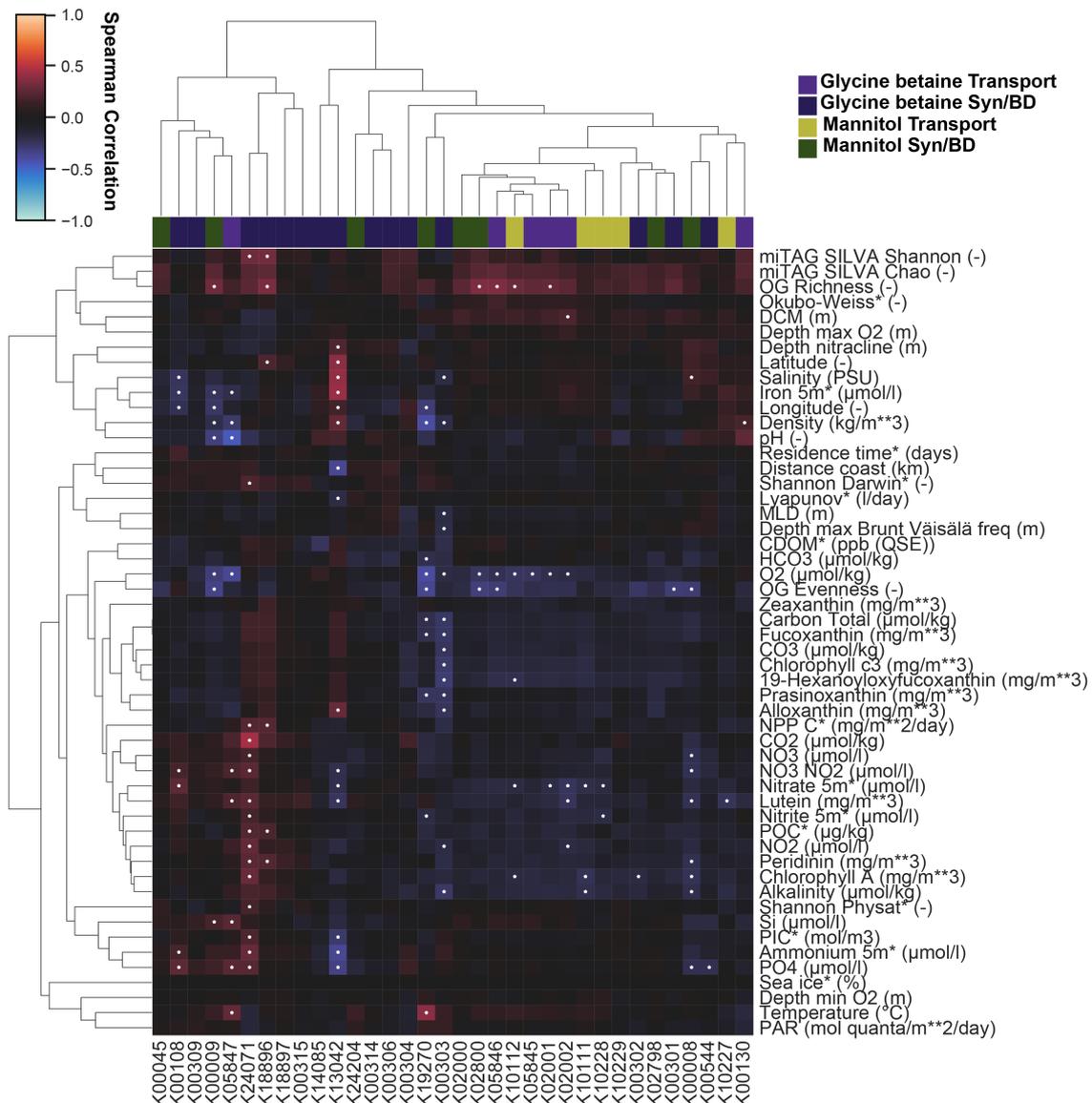


Figure S10. Spearman correlation against environmental parameters from Tara. The global abundance of key orthologs involved in the synthesis, breakdown, and transport of mannitol (green) and glycine betaine (purple) from eukaryotic metatranscriptomic data from Tara was correlated with core environmental parameters. The metatranscriptomic abundance (percent mapped reads) of each of the orthologs involved in the processing of the osmolytes shown here was assessed across the bacterial metatranscriptomic data from Tara Oceans using the MATOU-v1 dataset. The relative correlation of metatranscriptomic profiles for each of the orthologs considered was assessed by taking the Spearman's correlation of the sum of all orthologs associated with a KO at a given site and correlating it with available metadata. The Spearman's ρ is depicted with a heatmap that is clustered based on Bray-Curtis similarity. Significant correlations were determined with a Bonferroni multi-testing p-value correction, and significant correlations ($p < 0.001$) are depicted as a white dot.