



Reconciling Microbial Genomes and Water Quality Data through the length of the Chesapeake Bay

Keith Arora-Williams¹; Christopher Holder²; Eric Sakowski, PhD¹; Sarah Preheim, PhD¹

¹Environmental Health & Engineering, ²Earth and Planetary Sciences

Introduction

The USEPA defines water quality standards in order to protect animal habitat and economic productivity in the Chesapeake Bay (1). To ensure compliance, water quality indicator data are regularly collected throughout the Bay. These data show a trend of modest improvement, resulting from stricter management of wastewater treatment & air pollution (2). However, a significant fraction of the volume goes completely anoxic for much of the late summer.

Nitrogen (N) and Phosphorus (P) loading from many sources throughout the watershed spur phytoplankton growth. As their biomass drains to the bay, oxygen is respired by bacteria during its decomposition (3). In this work, we investigate how changes in water quality variables structure the abundances of genes and microbial genomes responsible for nutrient retention & cycling.

Figure 1. Sampling Station Locations: Note that some are nearly equal in latitudinal position. Most data was collected from CB3.3

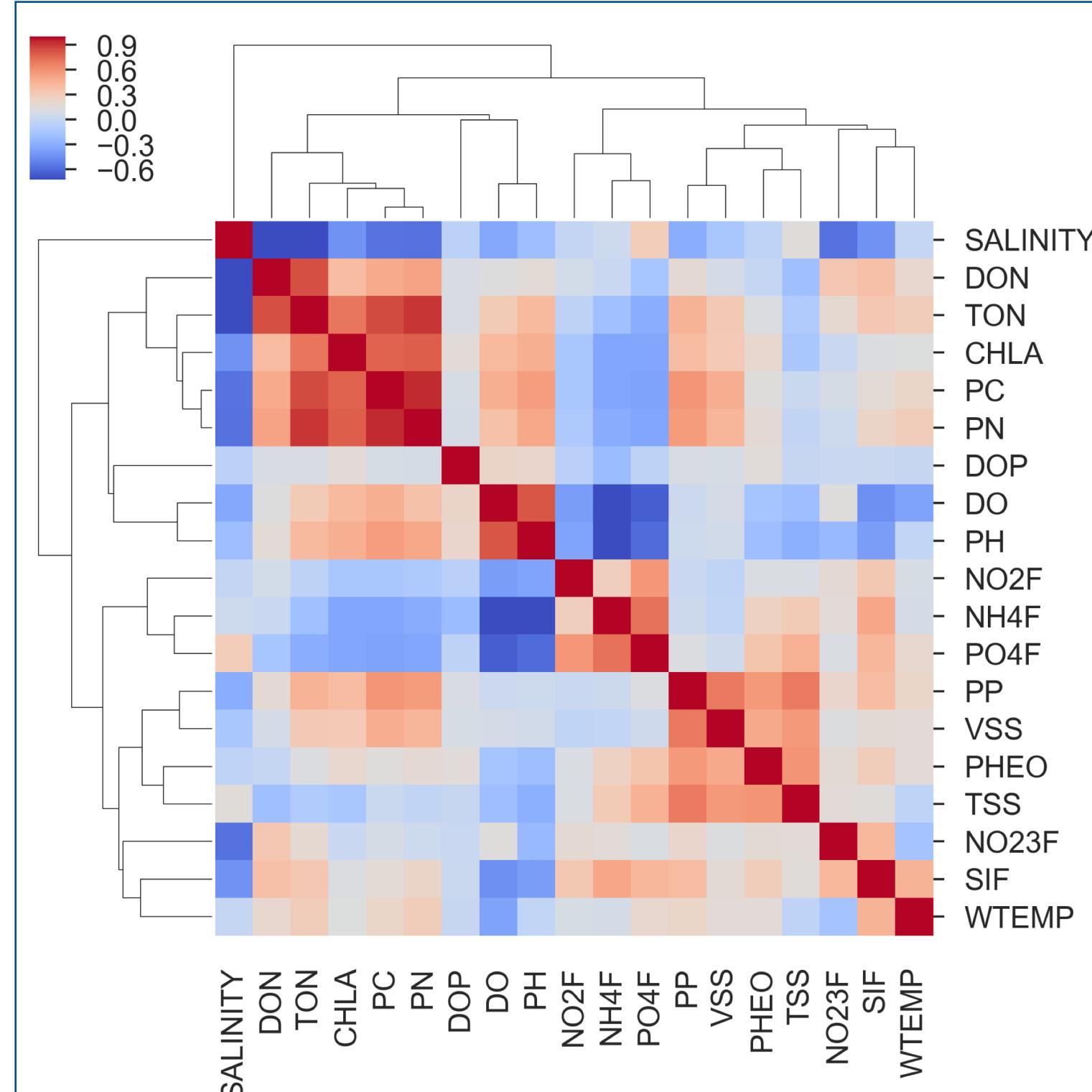
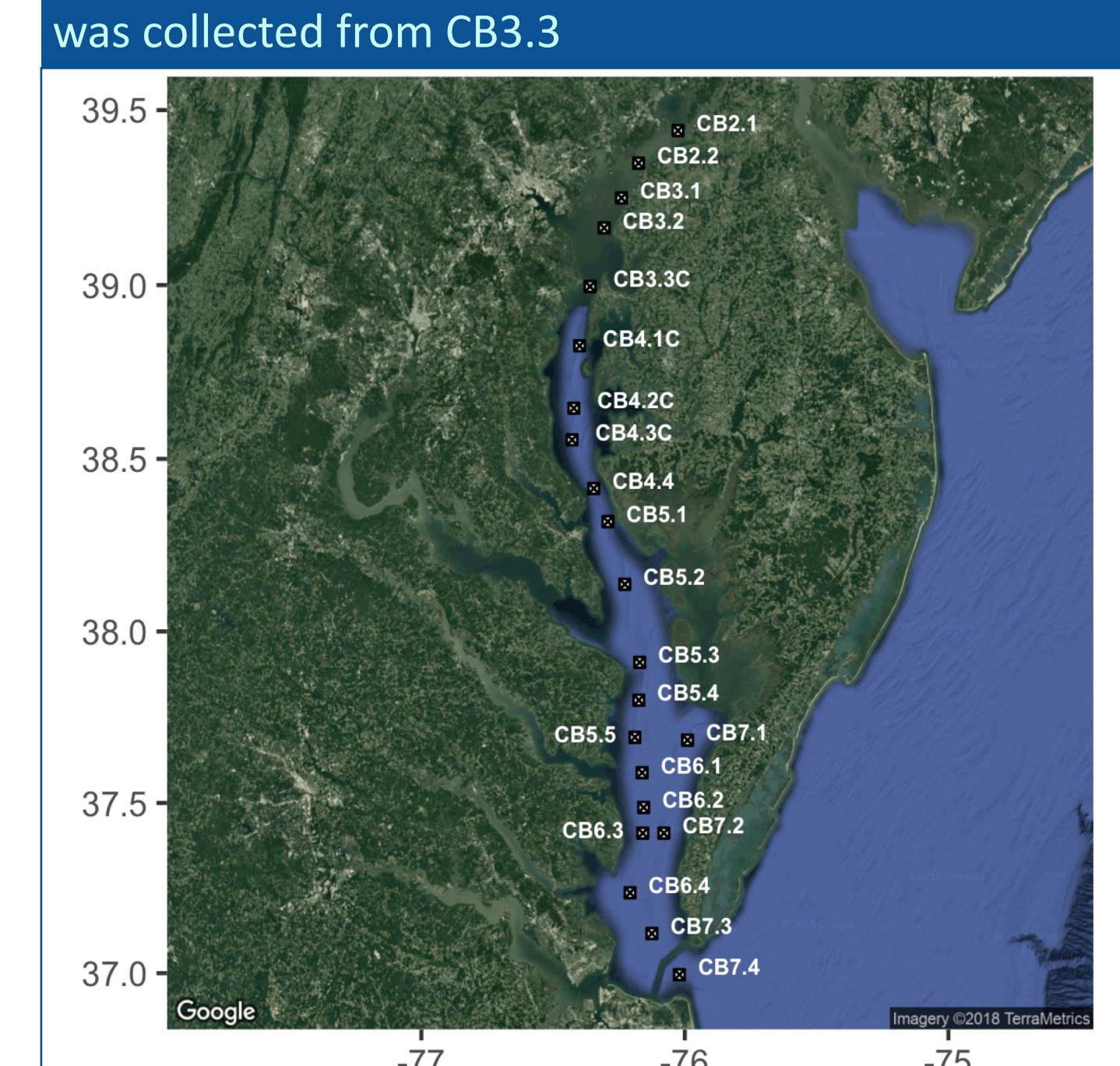
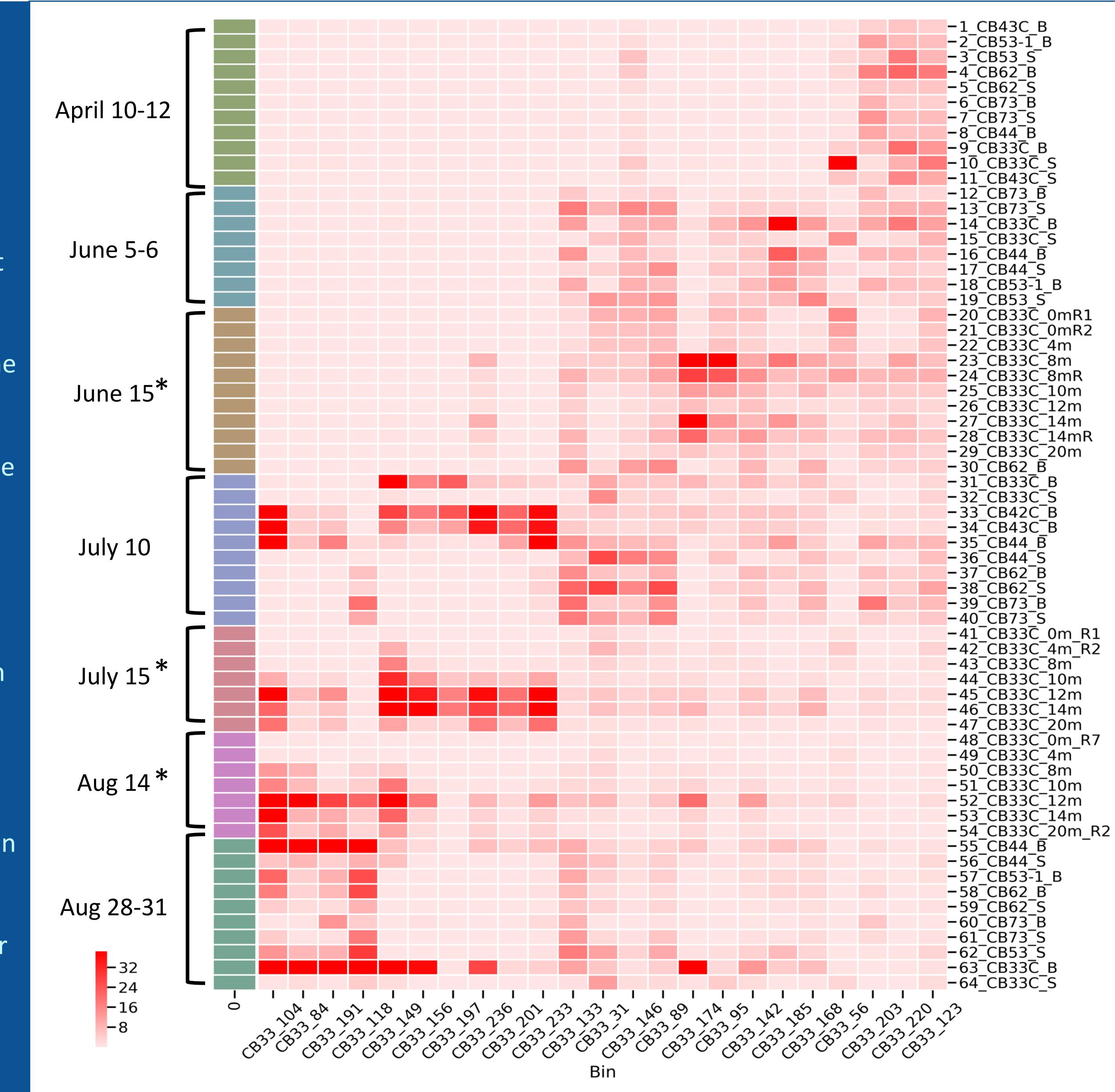


Figure 5. Genome Abundances: The heatmap shows the relative abundances of concordant genome bins (see Figs. 2 & 3). Colors in the leftmost column indicate the dates of sample collection in 2017. The asterisks indicate samples taken from different depths in the the CB33 set. The values shown are derived from the number of reads mapped to binned contig. Mapping from CB33 & CBrest bins were performed independently. Normalization was done to correct for bin size, assembly size, library size, and sequencing depth per sample group.



Methods and Materials

Water sampling

- Samples collected by Preheim lab, MD Dept of Natural Resources (DNR), & Old Dominion University (ODU) between April and August in 2017
- ODU & DNR sampled surface and bottom of stations in Fig. 1 ("CBrest" group)
- Preheim lab sampled multiple depths at CB3.3C ("CB33" group)
- Filtered at 0.22 µm pore size and DNA extracted from filters (Qiagen DNeasy PowerWater kit)
- Shotgun metagenomic libraries prepared using Nextera XT kit (Illumina)

Analysis

- Libraries assembled individually w/MEGAHIT (4), concatenated by sample group
- Consensus assemblies generated using BBMap and minimus2 (5) in series
- maxbin2 (6) and checkm (7) for genome binning and quality assessment
- Gene sequences were predicted using Prodigal (8)
- Proteins were annotated using the KEGG GhostKOALA tool (9)
- Water quality data from Chesapeake Bay Foundation DataHub (10)

Visualization

- Heatmap clusters were made using UPGMA method on Euclidean distances (11)
- t-distributed Stochastic Neighbor Embedded decomposition done using the Barnes-Hut method (learning rate = 200, perplexity = 30) (12).

Objectives

Data utilized in models designed to predict the size of the anoxic zone are accurate in the short term, under the current nutrient loading scenarios (13). To predict how they may respond to lower TMDLs and/or an altered climate, the following is required:

- Survey the microbial community composition and reconstruct the genomes of abundant populations in the Chesapeake Bay
- Use hydrodynamic/biogeochemical model to evaluate the role of dispersal in observed abundance patterns.
- Compare geo-specific deviations between observed and predicted water quality conditions with year-over-year shifts in local microbial communities

Figure 3. Homology between Independent Bin Sets: Scatter plot shows the proportion of each genome bin matched between data generated from independent sample sets. Only bins scored as >50% complete were compared. These show that only a few reconstructions are highly concordant.

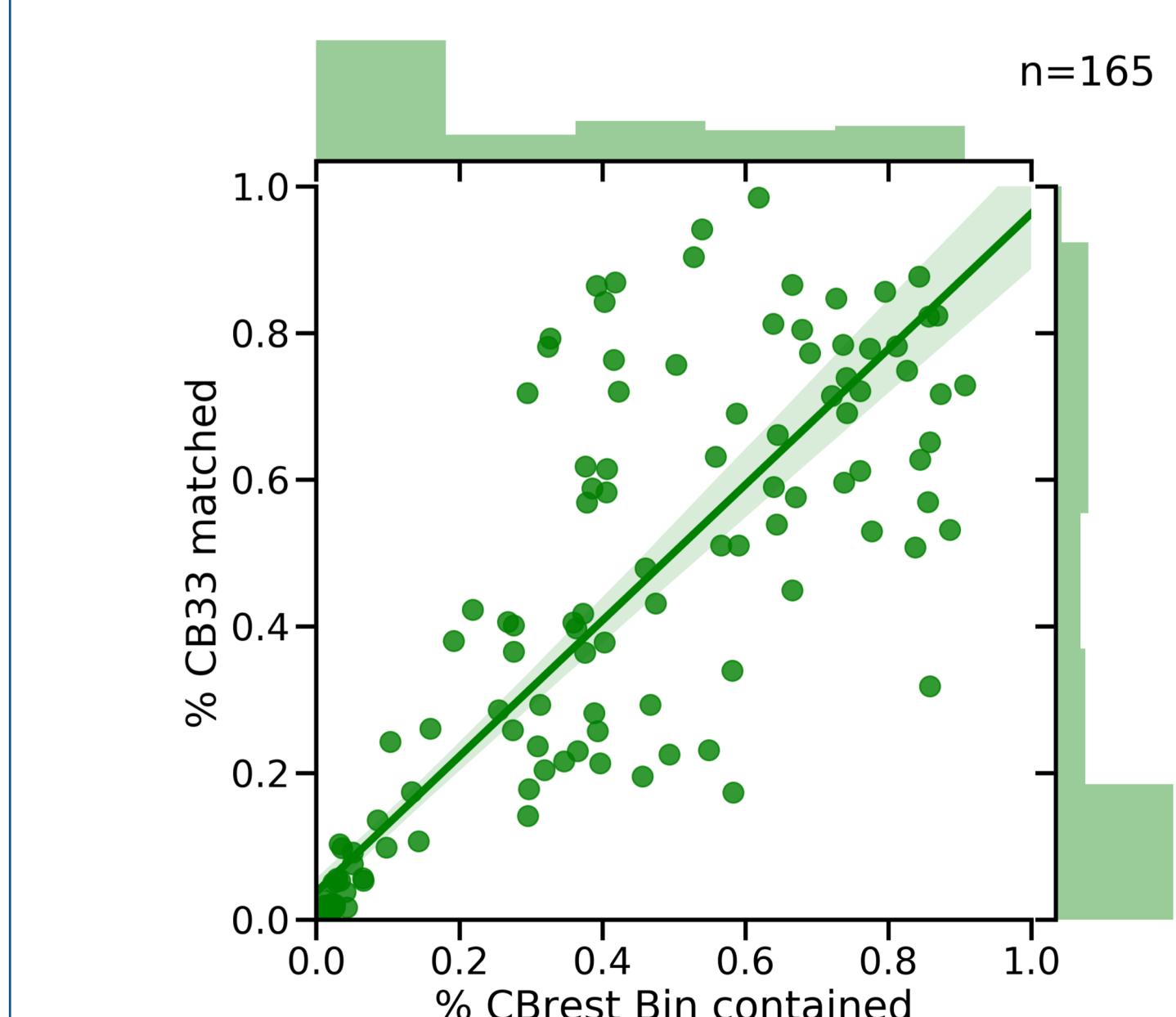
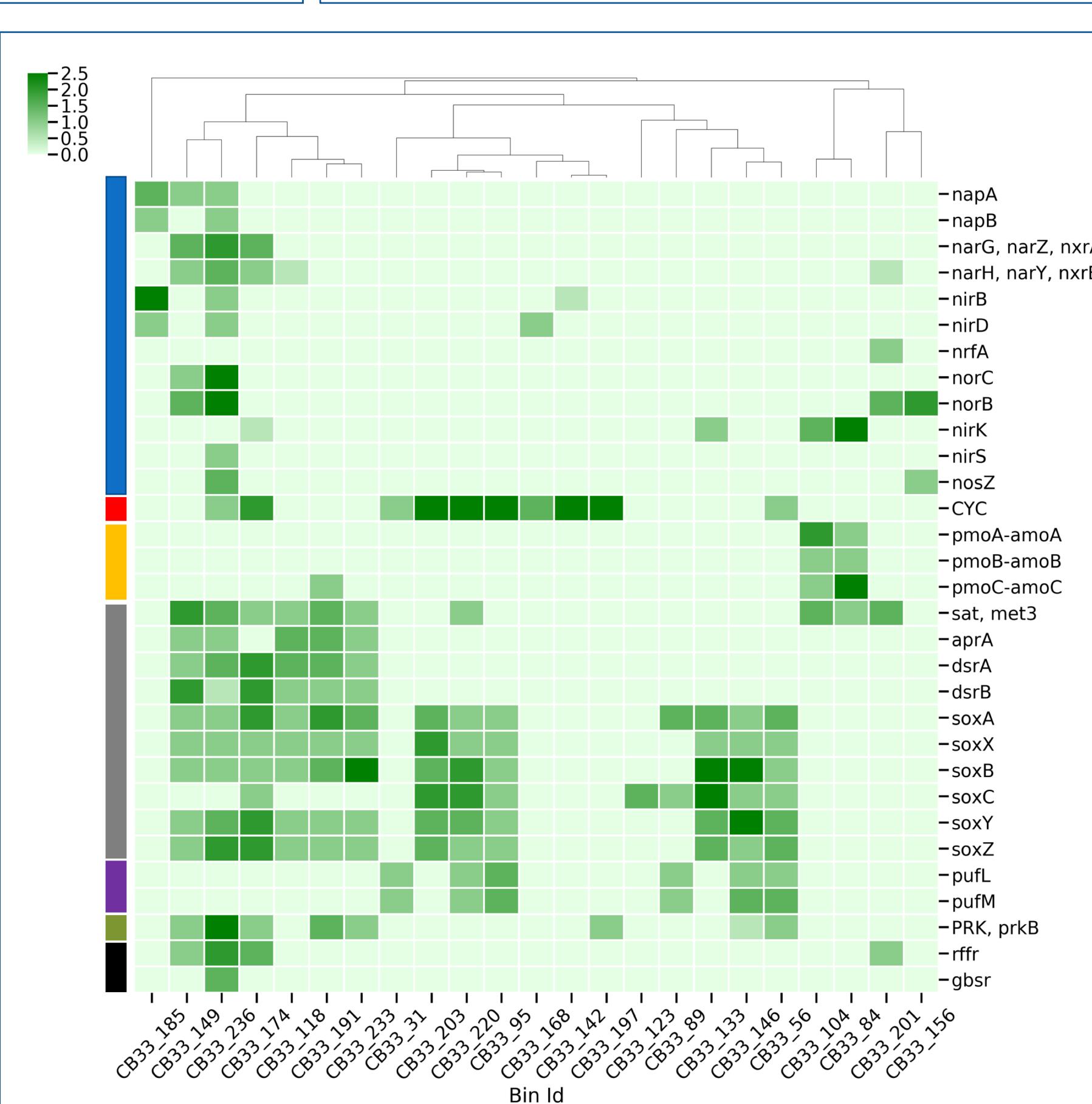


Figure 4. Marker Genes: The copy numbers of select marker genes (rows) within the top-most concordant bins (Fig 3), as well as correlation based clustering of bins (columns) are shown in the heatmap. The row of colors to the right indicate the biogeochemical process with which the gene is associated.

- Blue: Denitrification
- Red: Aerobic heterotrophy
- Orange: Methane / Ammonia Oxidation
- Grey: Sulfur Redox
- Purple: Anoxygenic photosynthesis
- Olive: Carbon Fixation
- Black: Iron reduction



Future Directions

- Use water quality data to determine the importance of differentially abundant genes and genomes.
- Link deviations between observed DO and ChesROMS model predictions to local microbial communities.
- Incorporate this data with 16S marker gene data covering multiple years and more sites to extrapolate genomic data across a more diverse set of conditions
- Mine publicly available gene expression data to establish whether patterns in gene co-occurrence can improve metabolic inference esp. for bidirectional enzymes (dsrB) or confounded annotations (pmoX-amoX)

Contact Information

Keith Arora-Williams
Preheim Lab | EHE Dept | JHU
Email: karoraw1@jhu.edu
Phone: (201) 421 - 1770

References

- USEPA. (2003) EPA 903-R-03-002
- Zhang, Q., et al. (2018). *Science of The Total Environment*, 637–638, 1617–1625.
- Bever, A. J., et al. (2018). *Journal of Geophysical Research: Oceans*, 123(9), 6392–6407.
- Li, D., et al. (2015). *Bioinformatics*, 31(10), 1674–1676.
- Sommer, D. D., et al. (2007). *BMC Bioinformatics*, 8(1), 64.
- Wu, Y. W., et al. (2014). *Microbiome*, 2(1), 26.
- Parks, D. H., et al. (2015). *Genome Research*, gr-186072.
- Hyatt, D., et al. (2010). *BMC Bioinformatics*, 11(1), 119.
- Kanehisa, M., et al. (2016). *Journal of Molecular Biology*, 428(4), 726–731.
- http://data.chesapeakebay.net/WaterQuality
- Podani, J. (2006). *Oikos*, 115(1), 179–185.
- Van Der Maaten, L. (2014). *The Journal of Machine Learning Research*, 15(1), 3221–3245.

Acknowledgements

Thanks to David Mohr from the GRCF for sequencing help; Drs Hugh Ellis, Ed Bouwer, Alan Stone, Scot Miller, and Ciaran Harman for their excellent courses; Gherman Uritskiy for help with technical issues; Capt. Ed Darwin for driving the boat; and Michael Lester, Denise Nowlin, and Jessica Elroy for administrative support.