

# Abundant and persistent sulfur-oxidizing microbial populations are responsive to hypoxia in the Chesapeake Bay

Keith Arora-Williams<sup>1</sup>, Christopher Holder,<sup>2</sup>  
Maeve Secor,<sup>1</sup> Hugh Ellis,<sup>1</sup> Meng Xia,<sup>3</sup>  
Anand Gnanadesikan<sup>2</sup> and Sarah P. Preheim<sup>1\*</sup>

<sup>1</sup>Department of Environmental Health and Engineering,  
Johns Hopkins University, 3400 N. Charles St.,  
Baltimore, MD 21218.

<sup>2</sup>Department of Earth and Planetary Sciences, 3400  
N. Charles Street, Baltimore, MD 21218.

<sup>3</sup>Department of Natural Sciences, University of Maryland  
Eastern Shore, Princess Anne, MD 21853.

## Summary

The number, size and severity of aquatic low-oxygen dead zones are increasing worldwide. Microbial processes in low-oxygen environments have important ecosystem-level consequences, such as denitrification, greenhouse gas production and acidification. To identify key microbial processes occurring in low-oxygen bottom waters of the Chesapeake Bay, we sequenced both 16S rRNA genes and shotgun metagenomic libraries to determine the identity, functional potential and spatiotemporal distribution of microbial populations in the water column. Unsupervised clustering algorithms grouped samples into three clusters using water chemistry or microbial communities, with extensive overlap of cluster composition between methods. Clusters were strongly differentiated by temperature, salinity and oxygen. Sulfur-oxidizing microorganisms were found to be enriched in the low-oxygen bottom water and predictive of hypoxic conditions. Metagenome-assembled genomes demonstrate that some of these sulfur-oxidizing populations are capable of partial denitrification and transcriptionally active in a prior study. These results suggest that microorganisms capable of oxidizing reduced sulfur compounds are a previously unidentified microbial indicator of low oxygen

in the Chesapeake Bay and reveal ties between the sulfur, nitrogen and oxygen cycles that could be important to capture when predicting the ecosystem response to remediation efforts or climate change.

## Introduction

Low-oxygen dead zones are expanding within estuarine and coastal ecosystems worldwide largely due to pollution and climate change (Breitburg *et al.*, 2018), making microbial processes active within dead zones increasingly important. Centuries of excessive nitrogen and phosphorous loading have contributed to the expansion of low-oxygen dead zones in coastal and estuarine ecosystems (Lotze *et al.*, 2006; Altieri and Gedan, 2015). Phytoplankton grow in response to nutrient pollution, which creates low-oxygen conditions when oxygen consumption during microbial decomposition of phytoplankton removes oxygen faster than it can be replenished. In the absence of oxygen, a unique suite of microbial processes become favourable, including those that influence greenhouse gas production, such as methane and nitrous oxide, change the mobility of toxins or nutrients, including arsenic and phosphate, and create reduced chemical species, like hydrogen sulfide and ammonia (Canfield *et al.*, 2005a). These processes may exacerbate habitat degradation or create other harmful products. Hydrogen sulfide, for example, is typically oxidized with oxygen either biotically or abiotically (Canfield *et al.*, 2005a). Hydrogen sulfide, as compared to decaying biomass, more easily diffuses out of the sediment, effectively transporting the location of oxygen demand upwards in the water column expanding the low-oxygen zone. However, other oxidation pathways exist, such as through anoxygenic photosynthesis or sulfide-oxidation coupled to denitrification, which could act to reduce the demand for oxygen. Thus, by identifying important microbial processes active within low-oxygen ecosystems, we hope to improve our understanding of the ecosystem consequences associated with larger and more common low-oxygen environments within estuarine ecosystems.

Received 13 July, 2021; revised 7 March, 2022; accepted 12 March, 2022. \*For correspondence. E-mail [sprehei1@jhu.edu](mailto:sprehei1@jhu.edu); Tel. (+410) 516 6632; Fax (+410) 516 8996.

An increasingly popular approach for studying uncultured microorganisms is to use non-targeted DNA and RNA sequence-based techniques to understand key biogeochemical processes (Reed *et al.*, 2014; Louca *et al.*, 2016b; Yabusaki *et al.*, 2017). These techniques have revealed the importance of novel or unknown microbial processes, such as the cryptic sulfur cycle that is difficult to detect based on chemical profiles alone (Reed *et al.*, 2014; Louca *et al.*, 2016b). The abundance of taxonomic or enzymatic markers, like the 16S ribosomal RNA (rRNA) or dissimilatory bisulfite reductase genes (*dsrAB*), can often change with environmental conditions in a way that suggests sequence-based information may be a useful quantitative geochemical biosensor (Smith *et al.*, 2015; Preheim *et al.*, 2016; Arora-Williams *et al.*, 2018). If so, observations of changes in microbial processes from 16S rRNA gene sequences and shotgun metagenomics could add an additional validation data set to models developed for management or remediation, such as within the Chesapeake Bay (Cerco and Noel, 2017). More work is needed to understand whether unknown or cryptic microbial processes occurring in estuarine ecosystems are important to consider from a modelling and management perspective.

The Chesapeake Bay is an important estuarine system with complex seasonal and long-term chemical dynamics (Kemp *et al.*, 2005). It is the largest estuary in North America and receives significant nutrient inputs from seven major rivers along its length. Periodic mixing by winds and tides stirs this freshwater into the interior, setting up a density gradient that drives an inflow of Atlantic water and outflow of fresher surface water. During the summertime, warming of the surface and lower winds significantly reduce the magnitude of such mixing events, greatly reducing the downward mixing of oxygen, allowing hypoxic waters to accumulate in the main centre channel or main stem of the Bay.

Previous investigations in the Chesapeake Bay have revealed that temperature and salinity are major drivers of microbial community composition, but more work is needed to determine the role of microbial metabolisms in biogeochemical processes occurring within low-oxygen environments. There is a strong and repeatable seasonal shift in microbial community composition in surface waters (Kan *et al.*, 2006; Wang *et al.*, 2020) and with depth associated with the transition to anoxic conditions (Crump *et al.*, 2007). At lower resolution, with methods such as denaturing gradient gel electrophoresis, no significant spatial resolution in community structure was observed (Kan *et al.*, 2006). Using the higher-resolution technique associated with high-throughput sequencing of 16S rRNA genes, the spatial pattern along the length of the Bay becomes the second most important driver of community composition (Wang *et al.*, 2020) after

seasonal changes. These investigations focused on microbial communities in surface waters, where oxygen levels remain relatively high. Investigations of the microbial response to low-oxygen environments at the bottom of the Chesapeake Bay have been restricted to lower-resolution techniques, such as denaturing gradient gel electrophoresis, with a limited amount of 16S rRNA gene sequencing (Crump *et al.*, 2007) and gene expression through metatranscriptomics analysis (Eggleston *et al.*, 2015). These studies revealed that shifts in diversity and microbial community structure occur with anoxic conditions (Crump *et al.*, 2007) and key microbial processes that differ between oxic and anoxic water samples (Hewson *et al.*, 2014), highlighting sulfate reduction as a key difference. Few studies have focused on microorganisms that oxidize reduced sulfur species in the Chesapeake Bay, even though these microorganisms are common in other low-oxygen coastal and estuarine ecosystems (Lavik *et al.*, 2009; Walsh *et al.*, 2009; Glaubitz *et al.*, 2013; Li *et al.*, 2021). Abiotic sulfide oxidation is the only sulfide-oxidizing process captured in important Bay models (Cerco and Noel, 2017). However, it has been argued that photosynthetic sulfide oxidation should outcompete abiotic processes under most conditions (Hanson *et al.*, 2013), and evidence from the Chesapeake Bay suggests a small population of phototrophic sulfide oxidizers, specifically *Chlorobi* (Findlay *et al.*, 2015), are sufficient to explain the oxidation of sulfide in the Bay.

This work aims to identify the factors that influence the spatial, temporal and depth distribution of microbial populations in the Chesapeake Bay, focusing on changes to the microbial community in response to oxygen concentrations less than  $2 \text{ mg l}^{-1}$ . We analysed microbial populations in the Bay with 16S rRNA gene sequences, metagenome-assembled genomes (MAGs) and metatranscriptomics, and compared them with changes in chemical parameters and transport of particles from a physical model of the Bay. Unsupervised cluster analysis demonstrates three main groups of samples based on either water quality or microbial community composition, with the greatest correspondence between them attributed to the selective effects of season, salinity and hypoxia. Predictions of hypoxic conditions and taxa enriched in bottom water consistently highlight taxa and functions associated with sulfide oxidation. MAGs of sulfide-oxidizing members of *Thiomicrospirales*, *Thioglobaceae* and *Sedimenticolaceae* families verify the presence of genes associated with autotrophic sulfide oxidation and in some MAGs, such as *Sedimenticolaceae*, the ability for at least partial denitrification. We demonstrate these MAGs are persistent and active populations through mapping metatranscriptomic reads from a previous study (Hewson *et al.*, 2014) to these MAGs. This

represents a substantial deviation from existing models of sulfide oxidation in the Chesapeake Bay and highlights important targets to better understand microbial metabolism in low-oxygen environments of the Chesapeake Bay, as an important example of the microbial response in critical estuarine ecosystems.

## Results

### *Site description and water conditions*

A total of 236 samples were used for 16S rRNA gene analysis, generated from samples taken throughout the Chesapeake Bay in July and August 2016 and from April to September 2017 (Fig 1 and Supporting Information Fig. S1). Of these samples, 61 were surface samples taken from 1 m depth in 2017, with fewer stations in the Upper Bay, whereas 175 were taken 1 m from the bottom from most stations in both 2016 and 2017. The Chesapeake Bay hypoxic zone, which refers to water with less than  $2 \text{ ml l}^{-1}$  oxygen, was qualitatively similar between 2016 and 2017 (Supporting Information Table S1 and Fig. S2) and average with respect to maximum daily volume, total annual hypoxic volume and duration compared to historical values from 1985 to 2019 (Friedrichs, 2019). Precipitation, which can partially explain inter-annual variability of hypoxic volume in the Bay (Zhou *et al.*, 2014), measured as either depth change or discharge was qualitatively similar in 2016 and 2017 (Supporting Information Fig. S3). Some measured chemical and environmental parameters strongly and significantly co-vary. Latitude strongly co-varies with total nitrogen, dissolved organic nitrogen, total phosphorous, pH (Supporting Information Fig. S4A–D) and salinity (Supporting Information Fig. S5A). Water temperature, light, precipitation and discharge change substantially with seasonality (Supporting Information Fig. S6). In addition, dissolved oxygen is correlated with pH and chlorophyll a, and anti-correlated with ammonia and phosphate (Supporting Information Fig. S7). Other strong associations in the chemical data set include salinity and nitrate (Supporting Information Fig. S5B) and total nitrogen and non-chlorophyll a photosynthetic pigments (pheophytin and phycocyanin; Supporting Information Fig. S5C,D).

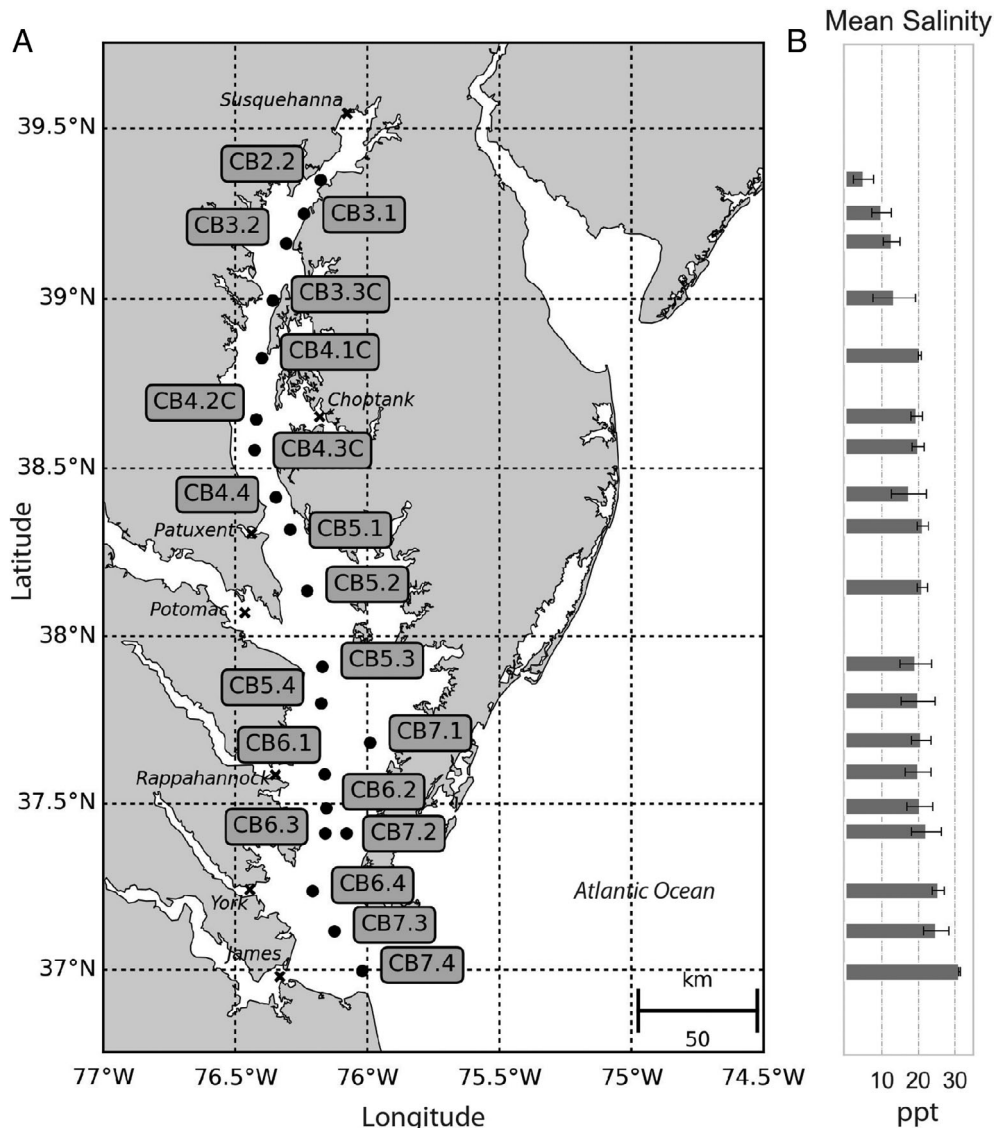
### *Variation in taxonomy in space and time*

The taxonomic distribution of the microbial community, assessed with 16S ribosomal RNA gene sequencing and analysed with differential abundance analysis, changed with season and with station but less with water column position from surface to bottom (Fig. 2 and Supporting Information Table S2). The V4 region of the 16S rRNA gene was sequenced resulting in over 20 million paired-

end reads across 236 samples after quality filtering. Since there are many known biases associated with 16S rRNA gene analysis (Bonk *et al.*, 2018) resulting in an imperfect representation of the in situ bacterial community, we focused on differences between samples, assuming biases are similar between samples. To determine significant differences in community membership in space and time, differential abundance analysis was repeated at genus, family, class, or order level for different sample categories. Approximately 100 genus, family, class, or order taxonomic groups were differentially distributed with season and position in the Bay, with a majority classified as Actinobacteria, Bacteroidia, Gammaproteobacteria and Alphaproteobacteria (Supporting Information File S1). In comparison, only 12 genus, family, class or order taxonomic groups were differentially distributed between the bottom and top of the water column when all samples were used (Supporting Information Table S3) and seven groups were identified when paired surface and bottom water samples were used (Supporting Information Table S4). Cyanobacteria, either Oxyphotobacteria or *Synechococcus* (Supporting Information Fig. S8A), were found to be more abundant in the surface samples ( $p < 0.01$ ). For the paired data set, bacteria involved in sulfur cycling were enriched in bottom water samples ( $p < 0.01$ ), one in the sulfate-reducing family Desulfobulbaceae (Supporting Information Fig. S8B) and an undefined genera in the order Chromatiales (Supporting Information Fig. S8C), which contains sulfide-oxidizing representatives.

### *Samples cluster differently based on water quality or microbial parameters*

To evaluate the factors that strongly influence community composition at the amplicon sequence variant (ASV) level, we used k-medoids clustering on Euclidean distances of the first three axes of decomposed data separately on microbial community composition and water quality data. The optimum number of clusters, identified using Hartigan's metric, was three for both microbial community composition and water quality data and these clusters captured a substantial amount of the variation between samples (34%). We refer to these clusters as habitats H.1–H.3 for groups of samples based on water quality measurements (Fig. 3), or microbial clusters M.1–M.3, for samples clustered by microbial community composition (Fig. 4). Using analysis of variance, we determined that the most significant drivers of the separation between water quality habitat clusters were variation in dissolved oxygen (DO) and total nitrogen (TN) levels, as well as salinity and pH (Supporting Information Table S5). H.1 contains only bottom samples from stations CB2.2 to CB5.3 (lower salinity), H.2 contains only bottom samples



**Fig. 1.** Distribution of sampling sites in the Chesapeake Bay and associated salinities.

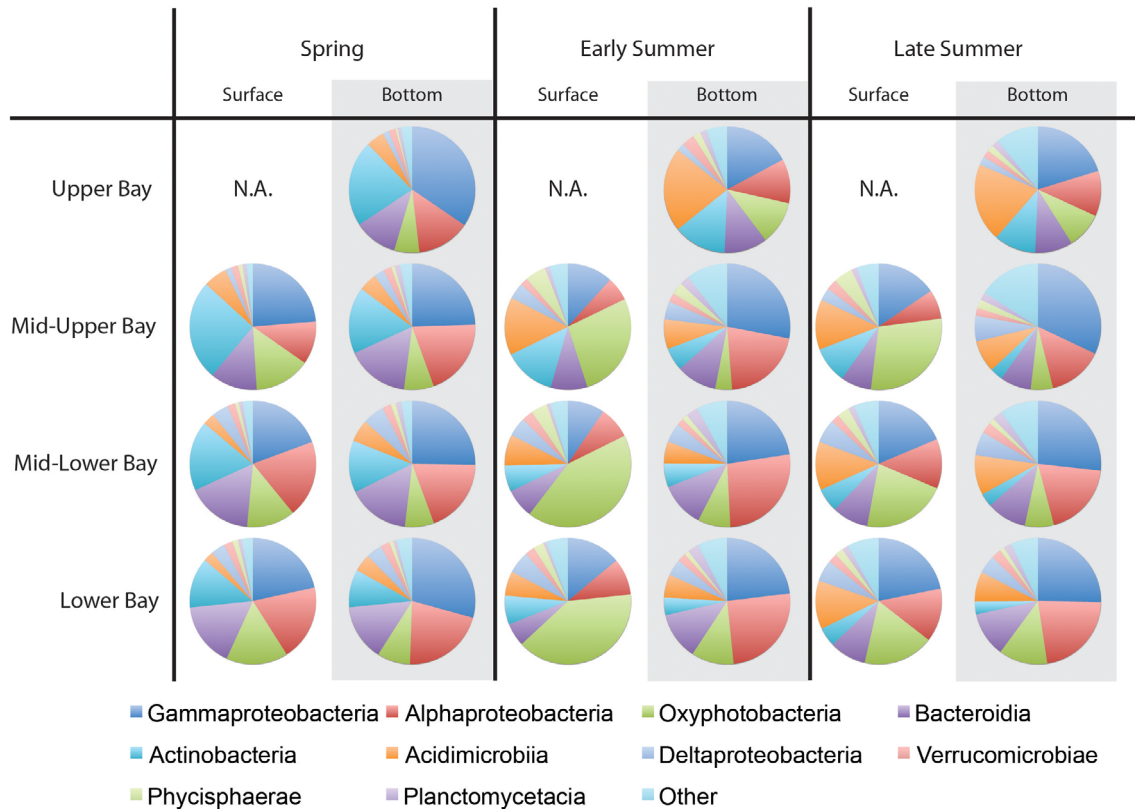
A. Map of the Chesapeake Bay with sampling sites (labelled with dots), and major tributaries (labelled with Xs)

B. Change in mean salinity with latitude along the Chesapeake Bay. Mean salinity (in ppt) shown with error bars indicating the standard deviation in salinity at each position in the Bay.

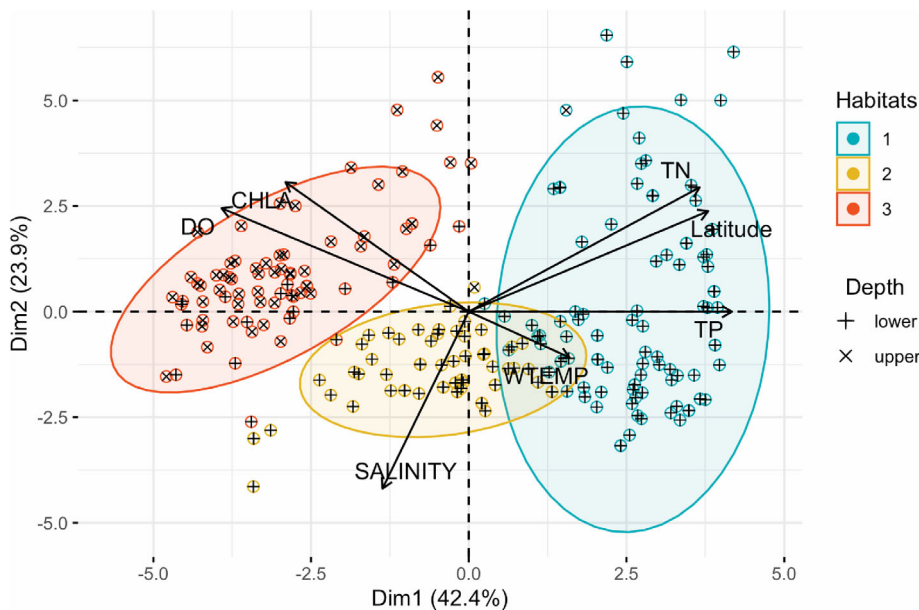
from stations CB5.1 to CB7.4 (higher salinity) and H.3 contains all surface samples and 17 bottom water samples from April and May regardless of salinity (Supporting Information Fig. S9A and C).

Although three clusters of samples were also observed when clustering by microbial community compositions, the two methods did not result in identical sample grouping, with only 64% of samples overlapping when comparing the most similar clusters (Supporting Information Table S6). Differences in environmental variable means between groups (Supporting Information Fig. S10) as well as sample clustering in space and time (Supporting Information Fig. S9) show microbial clusters were broken into a summer freshwater

cluster M.1, and a summer marine cluster M.2, and a spring cluster M.3. In general, microbial clusters were strongly influenced by season, but salinity drove the microbial clustering for summertime samples (Fig. 4). In microbial clusters, summertime surface samples were more commonly found in the same cluster as bottom water samples from the same part of the Bay, in contrast with water quality clusters, in which all surface samples clustered together regardless of location (Supporting Information Fig. S9). This difference was associated with the largest difference in sample clustering between microbial clusters and water quality habitats. These results suggest that factors influencing the microbial community composition may not be solely



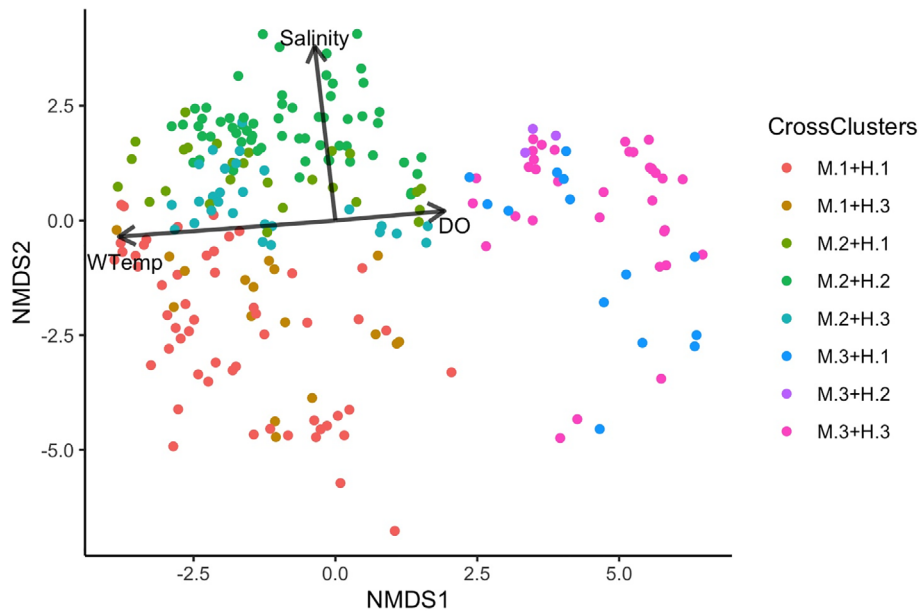
**Fig. 2.** Class-level taxonomic classification of prokaryotic community in the Chesapeake Bay in space and time from 2017. Samples were grouped based on time of sampling, water column position and station (see Supporting Information Table S2 for numbers of samples within each category). Other represents 38 taxa that were not shown.



**Fig. 3.** Clustering and biplot analysis of the factors that explain the sample grouping based on water quality data. Principal component analysis was used to visualize water samples clustered by differences in water quality parameters. This resulted in three clusters, represented by different colours, with the amount of variation shown in parentheses next to axis labels. Symbols indicate whether samples were collected from 1 m from the surface (upper) or 1 m from the bottom (lower) of the water column. Water quality parameters that explain the most of the variation between samples are projected onto the plot. WTEMP, water temperature; TN, total nitrogen; TP, total phosphorous; CHLA, Chlorophyll a; DO, dissolved oxygen.

determined by the variation in measured water quality parameters. The microbial community may be more sensitive to certain water quality parameters, such as salinity and

temperature, than others, or microbial community composition could be influenced by the movement of cells through mixing or sinking.



**Fig. 4.** Clustering based on microbial community composition and overlap with water quality clusters using non-metric multidimensional scaling (NMDS). Samples are clustered in space based phylogenetic isometric log-ratio transformation of the microbial community using the k-medoids algorithm on Euclidean distances. Samples are coloured according to their presence in one of the three microbial community-based clusters (M.1–M.3) plus their membership in one of the three water quality habitat clusters (H.1–H.3). M.1 + H.1 (red), M.2 + H.2 (green) and M.3 + H.3 (pink) have the highest number of overlapping samples. Samples with membership in other combinations of microbial clusters and water quality habitat cluster are coloured in various shades of browns, blues and purples. NMDS stress after projecting in two dimensions is low (0.15), suggesting clustering captures a majority of the variation across samples. Salinity, dissolved oxygen (DO) and water temperature (WTemp) are the major environmental factors. These results support the idea that the microbial community is strongly influenced by water quality parameters, but that overall water quality values explain about 55% of community composition.

#### *Particle back-tracking model provides insight into physical factors influencing microbial community assembly*

A particle back-tracking model (Mao and Xia, 2020) was paired with a hydrodynamic model of the Bay (Xia and Jiang, 2016) to determine how water movement influenced the microbial community composition at each sampling station. The model predicted the latitude and depth at which 300 inert particles were found 9 days prior to sample collection. We found that, rather than originating from the depth where they were collected, the particles collected at the surface and the bottom came from depths that substantially overlap (Supporting Information Fig. S11). On average, 50% of particles from paired surface and bottom water samples collected at the same station and time were predicted to be in the water column at the same depth interval 9 days prior to sampling. Particles came from different places in the Bay, as particles from the surface came from the north, while particles at the bottom came from the south (Supporting Information Fig. S12). This model has certain limitations making it an incomplete representation of the movement of microorganisms in this system, including that particles do not grow, decay or settle. In spite of these limitations, the results from this model suggest that relatively fewer

taxa were differentially distributed between surface and bottom water samples than season or position in the Bay because of the large amount of mixing of particles between surface and deep water. Dilution rates implied by this analysis (of order 4%–6% per day) are consistent with the evolution of salinity in the bottom water as it flows northward.

#### *Microbial community taxa and functions associated with sulfur metabolism are predictive of hypoxia*

Although temperature and salinity were found to be the most important drivers of microbial community composition, dissolved oxygen also differed substantially across microbial clusters. Identifying microbial taxa or functions most strongly associated with low-oxygen conditions could provide insight into key microbial processes occurring in low-oxygen environments and could serve as a useful biosensor for the state of the ecosystem. To do this, recursive feature elimination with random forest models were used to classify samples with less than  $2 \text{ mg l}^{-1}$  DO based on microbial features. Three abstractions of the data were tested including ASVs (Supporting Information File S3), taxonomic families (Supporting Information File S6) and potential functions predicted with

FAPROTAX (Supporting Information File S6). All three abstractions performed well with 10-fold cross validated accuracy scores of 95%, 95% and 90% for ASVs, taxonomic families and functions, respectively, which suggests that oxygen status of the sample can be predicted with high confidence based on changes in microbial community composition or predicted function.

Regardless of abstraction type, taxa and functions associated with sulfur metabolism were consistently among the top predictors of hypoxia. The most important single feature for function-based prediction was chloroplasts, which was negatively correlated to hypoxia (Supporting Information Table S7). Other important functional predictions positively associated with hypoxia were typical for low-oxygen environments, including thiosulfate and sulfur reduction, denitrification, anoxygenic photosynthesis and methanotrophy. Anoxygenic photoautotrophy from FAPROTAX inference was two orders of magnitude greater in overall relative abundance than other predicted functions and was nearly entirely based on predictions from sulfide-oxidizing Ectothiorhodospiraceae populations, within Gammaproteobacteria (Supporting Information Files S3 and S6). An ASV of Ectothiorhodospiraceae was also a predictor of hypoxia using the population level model (Supporting Information Table S8). Alphaproteobacteria and Gammaproteobacteria families known to be associated with the oxidation of reduced sulfur compounds made up 3/19 family-level (Supporting Information Table S9) predictions and 8/19 population-level predictions. These include families and ASVs classified as Sedimenticolaceae, Chromatiaceae and Thiomicrospiraceae within the Gammaproteobacteria and Rhodobacteraceae and Rhodospirillaceae within the Alphaproteobacteria.

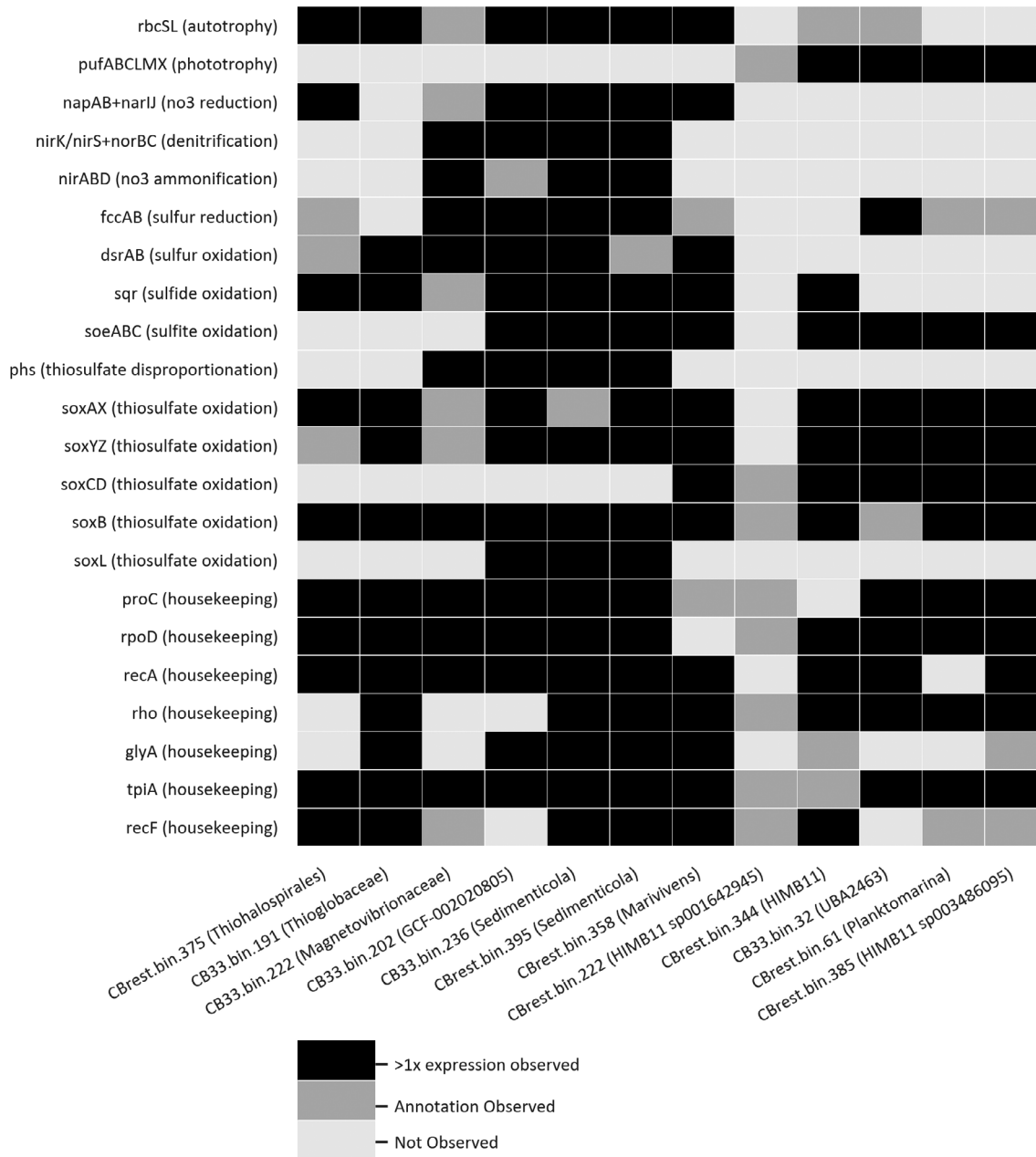
#### MAGs provide more insight into functions of sulfur-oxidizing organisms

To gain further insight into the metabolic capacities of abundant microbial populations, we generated MAGs from a total of 36 shotgun metagenomic data sets created from a diverse subset of samples from the mainstem used to make the 16S rRNA libraries (Supporting Information Fig. S1 and File S4). In addition, we added 16 shotgun metagenomic data sets created from samples collected from multiple depths at site CB3.3C in 2017, as this site is part of the mainstem sampling and experiences the longest period of hypoxic conditions. Shotgun metagenomic data were assembled and binned into MAGs, resulting in a total of 58 MAGs that passed our filtering criteria ( $\geq 80\%$  complete,  $\leq 10\%$  contamination). Our approach focused on medium quality draft MAGs or better (Bowers *et al.*, 2017), with special attention paid to those with higher per cent completion. Reads mapping to

MAGs were distributed over time and depth at one site (CB3.3 C; Supporting Information Fig. S13) or in surface and bottom samples from multiple sites (mainstem; Supporting Information Fig. S14) demonstrating these MAGs are representative of different types of organisms found in these samples.

Eight MAGs were classified as Sedimenticolaceae or Rhodobacteraceae, which match sulfur-oxidizing taxonomic groups predictive of hypoxia. Six out of these eight MAGs contained genes that support their role in the oxidation of reduced sulfur compounds (Fig. 5). All six had genes involved in thiosulfate oxidation through the sulfur oxidation (Sox) pathway (Ghosh and Dam, 2009), sulfite dehydrogenase (*soeABC*) and one or more of the following genes: oxidative forms of dissimilatory sulfite reductase (*dsrAB*) (Muller *et al.*, 2015), sulfide: quinone oxidoreductase (*sqr*), flavocytochrome c sulfide dehydrogenase (*fccAB*), photosynthetic reaction centre subunits (*puf* genes), ribulose-bisphosphate carboxylase (*rbc*). These genes are commonly involved with phototrophic or chemolithoautotrophic oxidation of reduced sulfur compounds (Ghosh and Dam, 2009). Only two other MAGs, classified as Phycisphaeraceae (CB33.bin.201 and CBrest.bin.370), had a taxonomic classification that matched a family predictive of hypoxia, although they did not harbour any key sulfur cycling genes.

Five additional MAGs, including those classified as taxonomic groups harbouring sulfur-oxidizing organisms, such as Thioglobaceae (Glaubitz *et al.*, 2013; Callbeck *et al.*, 2018), Magnetovibrionaceae (Bazylnski *et al.*, 2013) and Thiohalospirales (Sorokin *et al.*, 2008), also contained a substantial number of sulfur-oxidizing genes (Fig. 5). Interestingly, four MAGs (CB33.bin.202, CB33.bin.222, CB33.bin.236 and CBrest.bin.395) contained genes involved in denitrification, such as nitrate reductase (*nap*), nitrite reductase (NO-forming; *nirS*) and nitric oxide reductase (*nor*), although none contained genes to mediate the final step in denitrification (nitrous-oxide reductase, *nosZ*). These four MAGs were classified as Sedimenticolaceae (CB33.bin.202, CB33.bin.236 and CBrest.bin.395) and Magnetovibrionaceae (CB33.bin.222), which harbour taxa known to couple oxidation of reduced sulfur substrates to partial denitrification, including *Sedimenticola thiotaurini* SIP-G1 (Flood *et al.*, 2015) and *Magnetovibrio blakemorei* (Bazylnski *et al.*, 2013), respectively. To provide additional evidence that co-occurring nitrogen and sulfur cycling genes within MAGs are not the result of contamination, the phylogenetic relationship of key sulfur (*dsrA*, *soxX*) and nitrogen cycling (*nap*, *nir*) genes from these MAGs were compared to sulfur and denitrification genes from strains known to oxidize reduced sulfur species coupled with partial denitrification (Supporting Information Figs. S15–S20). A maximum likelihood tree was generated with these



**Fig. 5.** Metagenome assembled genomes (MAGs) containing genes mediating the oxidation of reduced sulfur species. MAG identity and their classification in parentheses on the x-axis and the gene designations with generalized role in parentheses are shown on the y-axis. Black boxes indicates genes that are both contained in the MAG and found with > 1X coverage in the 2010–2011 metatranscriptomic data. Dark grey boxes indicate the gene is found within the MAG, but not expressed, and light grey indicate the gene is missing from the MAG. Key genes associated with carbon fixation (*rbc*), anoxygenic photosynthesis (*puf*), oxidation of reduced sulfur species (*sox*, *soe*, *dsr*, *sqr*, *fcc*, *phs*), denitrification (*nap*, *nar*, *nir* *nor*) and a subset of housekeeping genes are displayed.

sequences plus the 50 most closely related sequences in the KEGG database. CB333.bin.236, CB333.bin.202 and CBrest.bin.395 cluster closely to *S. thiotaurini* SIP-G1 sequences for any of the genes that could be found in the MAGs. Only MAG CBrest.bin.395 contained all four genes. CB333.bin.222 clusters with *M. blakemorei* at *soxX*, *nirS* and *nap* loci but slightly more distantly at *dsrA*. It should be

noted that the three Alphaproteobacteria genomes, *M. blakemorei*, *Varunaivibrio sulfuroxidans* (Patwardhan and Vetriani, 2016) and *Magnetospira* sp. QH-2 (Ji *et al.*, 2014) cluster together at some, but not all loci, even though they are Alphaproteobacteria and couple sulfur oxidation with the reduction of nitrogen species. Nitrogen and sulfur cycling genes from MAGs clustering closely



with genes from known sulfide-oxidizing denitrifiers and MAG classifications associated with sulfide-oxidizing denitrifying taxa support the conclusion that microorganisms in the Chesapeake Bay have the potential to use partial denitrification in low-oxygen environments to oxidize reduced sulfur compounds.

#### *Expression of genes attributed to specific MAGs suggests highly active and persistent populations*

We wanted to know if and where genes found within the observed MAGs were expressed. We looked for expression of genes associated with MAGs within metatranscriptomic data from a study collected from mid-bay station CB 4.3C from April to October, 6–7 years prior to our sampling efforts (Hewson *et al.*, 2014). Water column samples ( $n = 18$ ) were collected in duplicate from above and below the pycnocline at various depths depending on in situ conditions to create metatranscriptomic libraries in this previous study, and those metatranscriptomic reads were mapped to MAGs from our data set with salmon (Patro *et al.*, 2017).

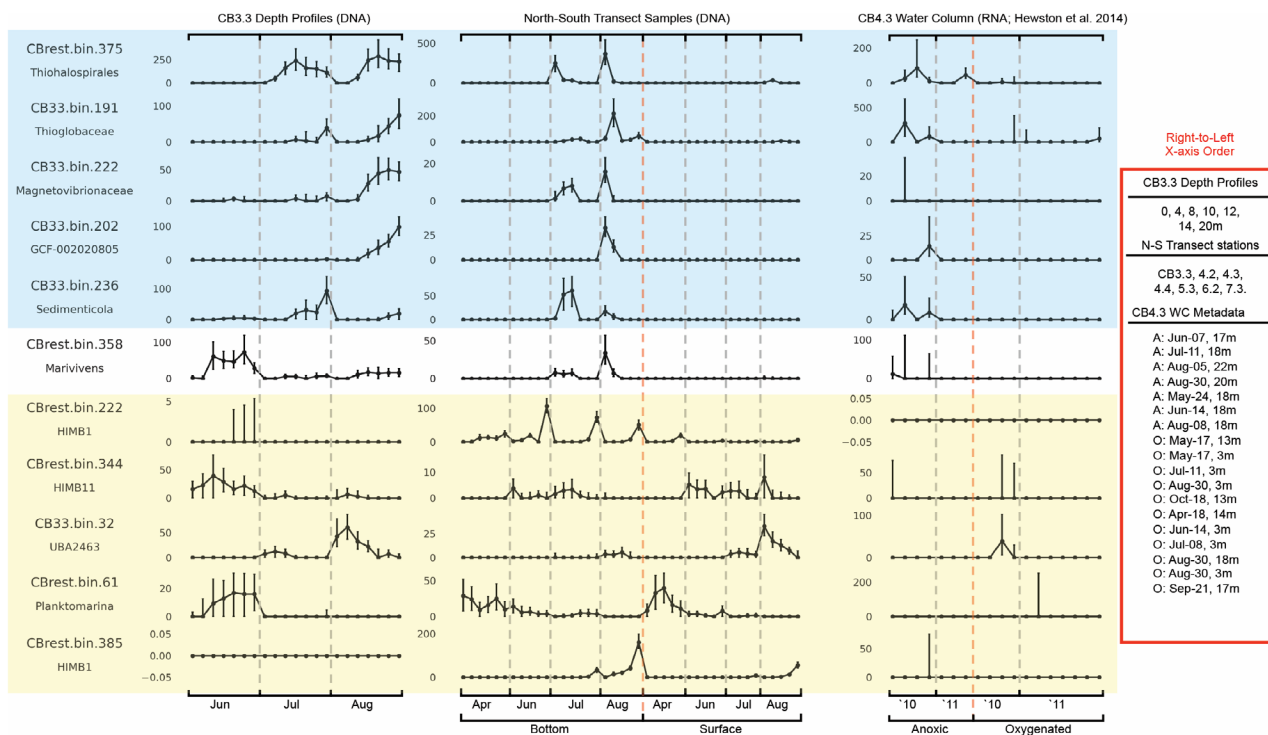
There is clear evidence that organisms represented by the genomes reconstructed in our work were present and transcriptionally active 6–7 years prior to our study (Fig. 6). For 21/58 MAGs, no expression was detected in any sample, which is not unexpected as all metatranscriptomes were taken from a single station. We observed transcripts from one or more samples mapping to an average of 1419 protein-coding genes across the remaining 37 MAGs, which include key genes involved in sulfur and nitrogen cycling genes (Supporting Information File S8). There are significant differences in the number of transcripts mapping to our MAGs per sample, with transcripts mapping to fewer than 11% of all genes within all MAGs in 11 samples, while the other 7 samples had transcripts map to between 18% and 54% of genes within MAG. Transcripts map to key genes involved carbon, nitrogen and sulfur cycling in at least 1 sample for 11 sulfide-oxidizing MAGs and 8 MAGs have transcripts mapping from samples within the anoxic water column (Fig. 6). In a few samples, notably in the anoxic water column on 30 August 2010 and throughout the water column on 11 July 2010, sulfur cycling genes are co-expressed with denitrification genes and carbon fixation genes from the same MAG (CBrest.bin.395), although the number of transcripts mapped to the denitrification genes tended to be lower than those mapped to key sulfur or carbon cycling genes (Supporting Information Fig. S21). Although not detected in MAGs, *nos* genes were found to have increased in expression during periods of anoxia as part of the previous analysis by Hewson and colleagues (2014). The results from mapping transcripts from the Hewson *et al.* data set to our

MAGs suggests that microorganisms capable of oxidation of reduced sulfur compounds in the Chesapeake Bay are persistent, abundant and active in low-oxygen environments and involved in at least partial denitrification.

#### **Discussion**

Our work provides insight into the microbial response to changing environmental conditions in the Chesapeake Bay estuary, focusing especially on microbial processes occurring in the deep, hypoxic water column. Overall, the microbial community composition in the Chesapeake Bay responds most strongly to variables that change with season and position in the Bay, such as temperature and salinity, although some microbial community members are more responsive to low-oxygen environments than others. Microorganisms involved in the oxidation of reduced sulfur species were consistently predictive of low-oxygen levels. Partial denitrification could allow these organisms to continue to metabolize when oxygen is low, although if and when nitrate is reduced completely to nitrogen gas by other microorganisms in the community harbouring the *nosZ* gene needs further clarification.

Sulfide oxidation, potentially coupled to denitrification, was not highlighted as a prominent response of the microbial community to low oxygen in the Chesapeake Bay before this study. Sequences derived from denaturing gradient gel electrophoresis bands using samples collected from anoxic waters were similar to known autotrophic sulfide-oxidizing and denitrifying microorganisms (Crump *et al.*, 2007), but the study focused on the succession in terminal electron acceptors for microbial respiratory processes (e.g. nitrate, iron and sulfate reduction), largely based on chemical observations. Previous work investigating differences in gene expression between surface and bottom oxic or anoxic water column samples highlighted sulfate reduction as an important water column process (Hewson *et al.*, 2014; Eggleston *et al.*, 2015). The top phylogenetic hit to the dissimilatory sulfur reduction database in the Eggleston and colleagues (2015) study (Table 3) was to the Gammaproteobacteria photosynthetic sulfide-oxidizing species *Allochrochromatium vinosum*. Since expression of genes involved in the oxidation of reduced sulfur species was not explicitly considered as part of their analysis, this raises the question of whether the thresholds for dissimilatory sulfur reduction were stringent enough to discriminate them from genes involved in sulfur oxidation (e.g. *dsrAB* which has both oxidative and reductive forms; Muller *et al.*, 2015). More work is needed to elucidate how the expression of genes involved in both sulfur oxidation and sulfate reduction change over the course of the season within the Chesapeake Bay. Phototrophic sulfide



**Fig. 6.** Metagenomic and metatranscriptomic read abundances mapping to metagenome assembled genomes (MAGs) with genetic potential for the oxidation of reduced sulfur compounds in the Chesapeake Bay. Shotgun metagenomic reads (*y*-axis; reads per million or RPM) mapping to MAGs (MAG names and classification on left) at one station (CB3.3) with depth, over multiple stations (North–South Transect) at the surface or bottom of the water column, and shotgun metatranscriptomic reads from one station (CB4.3) at various dates and depths. Error bars indicate standard deviation in RPM of different genes in the MAG. Legend on left indicates depths, stations and date/depth for each of the data sets mapped to the MAGs, ordered from right to left on the *x*-axis. Blue shaded regions indicate MAGs with chemolithoautotrophic potential and yellow shaded regions indicate MAGs with phototrophic potential.

oxidizers, such as *Chlorobi* in particular (Hanson *et al.*, 2013; Findlay *et al.*, 2015), were thought to be key in explaining how sulfide was being oxidized in anoxic waters (Luther *et al.*, 1988). None of the ASVs in our 16S rRNA gene data set (Supporting Information File S6) or any MAGs (Supporting Information File S4) were classified as *Chlorobi*, demonstrating these taxa were not abundant during our sampling period. However, they could still contribute to a substantial amount of sulfide oxidation, even at low population levels, as previously suggested (Findlay *et al.*, 2015). Our work highlights additional taxonomic groups that should be considered when studying sulfide oxidation in the Bay. Although both photosynthetic and non-photosynthetic sulfide oxidation were evaluated previously in the Chesapeake Bay, nitrate-driven sulfide oxidation was overlooked as a potentially important process in experimental incubations (Findlay *et al.*, 2015) and models (Findlay *et al.*, 2017). Our work suggests a re-evaluation of where and how sulfide is oxidized in the Chesapeake Bay and suggests additional work is needed to understand these rates and potential connections among the sulfur, oxygen and nitrogen cycles.

As in the Chesapeake Bay, sulfide-oxidizing microorganisms, including those with denitrifying capabilities, are abundant and active in the water column of low-oxygen estuarine or coastal ocean sites. These include the Baltic Sea, Saanich Inlet, Namibian coast and some parts of the South China Sea. Cell counts based on phylogenetic probes targeting the SUP05 cluster demonstrated that these sulfide-oxidizing denitrifiers can account for 10%–13% of cells in the Baltic Sea (Glaubitz *et al.*, 2013) and off the coast of Namibia (Lavik *et al.*, 2009). Similar to this study, sulfide-oxidizing taxa made up a large fraction of 16S rRNA gene sequence data in Saanich Inlet (Walsh *et al.*, 2009) where genomes of abundant sulfide oxidizers were reconstructed with partial denitrification capabilities, but were missing the *nosZ* gene (Louca *et al.*, 2016b). The genomes of sulfide oxidizers, some with the ability to denitrify, could also be reconstructed from metagenomic libraries from the South China Sea (Li *et al.*, 2021), although *Campylobacter* MAGs were found in addition to Gammaproteobacteria and Alphaproteobacteria MAGs. Although our study found sulfide oxidizers of Alphaproteobacteria and Gammaproteobacteria taxa

associated with hypoxia, Epsilonproteobacteria were also associated with hypoxia in the South China Sea (Laas *et al.*, 2016; Broman *et al.*, 2017; He *et al.*, 2020) and the Baltic Sea (Takai *et al.*, 2006). Evidence of abundant and persistent sulfide oxidation is not found across all low-oxygen estuarine and coast environments. For example, while sulfide-oxidizing organisms were found in sediments (Broman *et al.*, 2017) and deep stratified parts of the water column (He *et al.*, 2020), ammonia oxidation, including anaerobic ammonia oxidation, was a commonly identified microbial process associated with low oxygen in the South China Sea (Han *et al.*, 2022). One proteomics study in the Gulf of St. Lawrence also suggested that anammox bacteria were abundant at lower depths, which tend to have low oxygen. In addition, some mixotrophic organisms similar to SAR324 and ARCTIC96BD-19 with genes for the oxidation of reduced sulfur compounds were also enriched in deep waters (Colatriano *et al.*, 2015). However, this study did not focus on the deepest part of the Laurentian Channel where hypoxia is most persistent (Fennel and Testa, 2019). The Chesapeake Bay is not unique in harbouring abundant and persistent sulfide-oxidizing taxa with some capability for partial denitrification, and differences in the duration or extent of low-oxygen conditions (Fennel and Testa, 2019) likely influence when and where these organisms are found.

Results from metagenome reconstruction and the particle-tracking model suggest that movement within the water column could select for versatile microbes that tolerate a wide range of environmental conditions. Cells of *M. blakemorei* (Bazylinski *et al.*, 2013) and *S. thioaurini* (Flood *et al.*, 2015) contain sulfur inclusions theorized to provide metabolic flexibility under fluctuating environmental conditions (Dahl and Prange, 2006). Future studies could investigate whether cells in the Chesapeake Bay harbour intracellular sulfur inclusions, similar to their closest known cultured representative, and if so, how this influences metabolic flexibility as cells are mixed between depths. Genes in the SOX pathway, involved in the oxidation of thiosulfate, were found in sulfur-oxidizing MAGs. Thiosulfate is a stable product of abiotic sulfide oxidation (Canfield *et al.*, 2005b), so it suggests a substantial amount of abiotic sulfide oxidation that may occur when surface and bottom water is mixed. Sulfide oxidation in the Chesapeake Bay is thought to be a combination of abiotic and biotic sulfide oxidation (Findlay *et al.*, 2017), and future studies should investigate the source and fate of thiosulfate within this system. Our study is largely based on samples collected from the top and bottom of the water column throughout the Bay. More work is needed to determine whether different set of microbial populations thrive at mid-depths, especially near the oxycline where the oxidation of reduced substrates is

expected to be maximal and whether these cells are alive or active as they move within the water column.

The Chesapeake Bay estuary is an important example of a critical ecosystem at the terrestrial-aquatic interface. Our work is consistent with previous studies that found season and position within the Bay, which correlates with salinity, to be strong drivers of microbial community structure through cluster analysis. We found that these drivers of community structure are stronger than differences between surface and bottom water samples, possibly related to a large amount of mixing. Our strategy of using population level analysis with 16S rRNA gene sequencing and MAGs, paired with gene expression, revealed the previously unappreciated potential that oxidation of reduced sulfur species can be an important microbial community response to low-oxygen environments in the Chesapeake Bay that warrants further investigation. The abundance of sulfur oxidizers in low-oxygen environments may be facilitated by the capability for at least partial denitrification. As the biogeochemical consequences of sulfur oxidation via different processes may have important impacts on oxygen concentrations, nitrogen availability and pH (Cai *et al.*, 2017), more work is needed to determine how this omission could change predictions of the ecosystem response under future climate and pollution scenarios.

## Experimental procedures

### Sample collection for metagenomic analysis

A total of 247 surface and bottom water samples were collected by the Maryland Department of Natural Resources and Old Dominion University in conjunction with routine physicochemical profiles taken as part of the Chesapeake Bay Water Quality Monitoring Program (Zhang *et al.*, 2018). Samples were collected over the course of 11 cruises in July and August of 2016, and April through September in 2017 (Supporting Information Fig. S1). The full set of 20 sampling locations span the length of the Bay (Fig. 1). Water samples were collected from 1 m below the surface and from 1 m above bottom (variable based on total water column depth at each station). Replicate water samples were collected from a subset of stations and depths. Water for shotgun metagenomic and 16S rRNA gene library analysis was pumped into a 50 ml conical tube, placed immediately in a freezer on the boat, stored at  $-20^{\circ}\text{C}$  for up to 3 months and stored at  $-80^{\circ}\text{C}$  for up to 1 year before processing.

To augment the 16S rRNA gene and shotgun metagenomic analysis, samples collected from station CB3.3C in June, July and August of 2016 and 2017 were included in the analysis. Because these samples were

collected separately from the regular water quality monitoring program, associated physiochemical measurements are not identical, thus they were not included in 16S rRNA gene analysis. However, a subset of these samples was used in 16S rRNA replicate analysis to compare the impact of freezing unfiltered water as opposed to immediately filtering water at the time of collection (Supporting Information Fig. S22 and File S7). Fifty microliters of samples were collected from 0, 1, 2, 4, 6, 8, 10, 12, 14, 16 and 20 m depth on each date with a peristaltic pump, either immediately filtered onto 25 mm diameter polyethersulfone filter membranes with 0.22  $\mu\text{m}$  pore size (MilliporeSigma, Burlington, MA) or placed unfiltered into a 50 ml conical tube and placed onto dry ice. Two volumes of water were allowed to pass through the tubing after repositioning at a new depth before samples were collected. Samples were stored at  $-80^{\circ}\text{C}$  for up to a year before processing.

#### Environmental and water quality data

Paired laboratory measurements, routinely published as a part of the Water Quality Monitoring Program, of phosphate, chlorophyll a, nitrate, nitrite, ammonia, pheophytin, particulate carbon, total phosphorus, total nitrogen, dissolved organic nitrogen and dissolved organic phosphorus, as well as sensor data for dissolved oxygen, pH and salinity (Supporting Information File S2) were obtained from the Chesapeake Bay Foundation Data Hub (<https://datahub.chesapeakebay.net/>, last accessed date 1 September 2019). To determine the extent of inter-annual and seasonal effects, time was represented linearly as the number of days (d) elapsed since 20 December 2015 divided by 365 and 'seasonality' as  $-\sin(2\pi d/365)$ , peaking at the vernal and autumnal equinoxes, previously shown to be applicable to analysis of microbial communities (Gilbert *et al.*, 2012).

#### Particle tracking model simulations

To better understand how the hydrology of the Bay influences the microbial community composition, we simulated the movement of bacteria over time as neutrally buoyant particles using a hydrodynamic model of the Bay and meteorological conditions at the time of sample collection. An existing simulation using the Chesapeake Bay Finite Volume Coastal Ocean Model (FVCOM; Xia and Jiang, 2016) with 11 vertical levels and a horizontal grid with resolution varying from 270.9 m to 20.9 km was used to drive a series of particle-tracking model simulations (Hu *et al.*, 2010; Mao and Xia, 2020). These model simulations were designed and implemented to back track where the microbiological particles were originally from. Hourly Eulerian current velocity at the

computational elements from FVCOM is interpolated to the particle locations in the particle-tracking model, in which time intervals for the scheme resolution and horizontal random walk are set at 120 and 6 s, respectively. The two-dimensional particle-tracking model with no vertical displacement solves the nonlinear differential equation:

$$\frac{d\vec{x}}{dt} = \vec{v}(\vec{x}(t), t),$$

where  $\vec{x}$  and  $\vec{v}(\vec{x}, t)$  are the particle location and velocity vector at time  $t$ , respectively, with an explicit fourth-order Runge–Kutta multi-step method. Analysis was done on 300 neutrally buoyant particles tracked for 9 days prior to sample collection. For stations and sampling dates with both surface and bottom water samples, the overlap in depth distribution from particle locations predicted from this simulation was determined from the sum of the minimum count within each 2 m segment from 0 to 32 m of both surface or bottom samples divided by 300, which is the total number of particles tracked in each sample, in R (R Core Team, 2015) using hist, sum and min functions.

#### 16S rRNA gene and shotgun metagenomic library preparation

Fifty microlitres of frozen water were thawed in a  $32^{\circ}\text{C}$  water bath for 5 min and filtered onto 25 mm diameter polyethersulfone filter membranes with 0.22  $\mu\text{m}$  pore size (MilliporeSigma, Burlington, MA). DNA was extracted from the filter with the Qiagen PowerWater kit following the manufacturer's instructions, including the optional step of heating for  $65^{\circ}\text{C}$  for 10 min, prior to homogenization. Extracts were quantified using the Qubit dsDNA HS assay (ThermoFisher, Waltham, MA) and diluted to  $1.2 \text{ ng } \mu\text{l}^{-1}$  prior to a polymerase chain reaction (PCR) protocol using the U515 and E786 primers, both of which have been described previously (Preheim *et al.*, 2013). Briefly, a portion of the V4 region of the 16S rRNA gene was amplified in two consecutive reactions. The first reaction added the Illumina sequencing primer region to one end of the amplicon over 19–22 cycles and was performed in quadruplicate. The second reaction added an eight-base barcode over nine additional cycles and was performed in quadruplicate. Phusion High Fidelity Polymerase (New England Biolabs, Ipswich, MA) was used for PCR, and SYBR Green I (ThermoFisher) was added during quantitative PCR (qPCR). Amplicon libraries were sequenced on a MiSeq using 250 or 300 base pair, paired-end reads at the Genetic Resources Core Facility (Johns Hopkins University, School of Medicine, Baltimore, MD).

Shotgun metagenomic libraries were prepared from the same DNA extracts for a subset of main stem samples and the CB3.3C-only data set ( $n = 51$ ; Supporting Information Fig. S1 and File S2) using both the Nexterra XT DNA Library and Illumina DNA Prep Kits (Illumina Inc, San Diego, USA). Thirty-six samples were chosen from the 2017 main stem samples to span the range of environmental conditions, time and sites within the data set. Fifteen samples from the CB3.3C-only data set were chosen to represent consistently sampled depths for the three dates collected in 2017. Three samples from CB3.3C and one from the main stem were sequenced twice to determine reproducibility. Library quality and fragment size range was assessed on the Bioanalyzer DNA1000 (Agilent, Santa Clara, CA) and quantified with qPCR using the KAPA Library Quantification Kits (Roche Holding AG, Basel, Switzerland). Shotgun libraries were sequenced on an Illumina HiSeq 2500, resulting in a yield of between 4.4 and 17.6 million bp paired-end read pairs per sample. Raw sequences for the 16S rRNA gene and shotgun metagenomic libraries were deposited under NCBI's BioProject accession number PRJNA612045. Previously described metatranscriptomes collected from below and above the pycnocline near station CB4.3C during the summers of 2010 and 2011 were used to assess the gene expression patterns and longevity of populations observed in our samples (Hewson *et al.*, 2014). Data were downloaded from NCBI Sequence Read Archive (Bioproject PRJNA222777) or provided by the authors upon request.

#### 16S rRNA gene quality control, normalization and analysis

For 16S rRNA gene sequence analysis, demultiplexing was done through *filterbyname.sh* command in the BBMap software (Bushnell, 2014). Trimming, error filtering, chimera removal, ASVs and taxonomic assignment were performed using the DADA2 (v1.4) R package (Callahan *et al.*, 2016). Twenty negative and positive control samples were included in the analysis. ASVs were filtered out if they had a higher log median abundance among controls than environmental samples or were found only in one sample. Nine samples with fewer than 100 ASVs and/or 3000 total counts were discarded, resulting in a total of 236 samples (Supporting Information Table S10 and File S3). Functional annotations were provided by FAPROTAX (Louca *et al.*, 2016a) using taxonomic annotations assigned using SILVA SSU database (release 132). ASV sequences were aligned using the *cmalign* command of the Infernal software against the small subunit RFAM (RF00177) covariance model. In an approach described previously (Bowman and Ducklow, 2015), RAXML v8.2 was used to generate

a maximum likelihood tree assuming the GTRGAMMA model of evolution and the '-f' option to select a root that best balances subtree lengths.

#### Shotgun metagenomic analysis

Shotgun metagenomic sequence data were processed as follows. Poor-quality and human-derived sequences were removed from shotgun metagenomic sequence data using TrimGalore v0.6.5 and bmtagger, respectively. We used an approach in which individual samples were assembled separately first, then combined before gene calling and MAG generation, similar to the approach used by other studies (Hugerth *et al.*, 2015; Tully *et al.*, 2018). We also divided our combined assemblies into two groups, CB33 and CBRest, which was less computationally intensive than a co-assembly of data from all samples. Error-corrected reads were assembled individually by sample using megahit v1.2.9 using software defaults (Li *et al.*, 2015). Assemblies from each sample within co-assembly groups were concatenated to form two combined assemblies used for binning. One combined assembly included 16 samples from only CB3.3C and the other combined assembly included 39 total samples, some of which were also from site CB3.3C (Supporting Information File S2). Each combined assembly was deduplicated using the 'dedupe' command of bbmap v38.86 (Bushnell, 2014) using a minimum identity threshold of 97% and a maximum edit threshold of 50. The deduplicated assemblies were merged using the overlap-layout-consensus assembler minimus2 in the AMOS v3.10 software package with default settings except using an elevated minimum overlap length 40 bp (Treangen *et al.*, 2011). The resulting contigs were binned using MaxBin2 (Wu *et al.*, 2016), using coverage values from all sample data sets within the co-assembly group generated by Burrows-Wheeler Aligner (BWA, Vasimuddin *et al.*, 2019). Contig-spanning coverage was calculated per sample using BWA and bedtools and normalized by contig length. The arithmetic average of the normalized coverage of all contigs was taken per MAG, following the methods implemented in the binning pipeline metaWRAP (Uritskiy *et al.*, 2018). The MAG abundances were represented as the mean of the abundance of constituent contigs unless otherwise specified and samples with fewer than 1 million reads were not displayed. MAGs with similar taxonomic classifications and read mapping patterns between co-assembly groups were compared but found to contain dozens to hundreds of unique genes, so all MAGs were included in the final analysis.

Gene calling and initial gene annotation was done using prodigal (Hyatt *et al.*, 2010) and prokka (Seemann, 2014), respectively. Models used to

complement initial gene annotation were taken from the kofam software (Aramaki *et al.*, 2020), as well as select collection of sulfur-relevant models taken from the protein family database (PFAM). Searches were executed using HMMER3 (Eddy, 2011) and low scoring hits were filtered using either the predefined cut-offs specified by the kofam software or a minimum *e*-value of  $10^{-15}$  and a minimum query coverage of 0.35. Genes annotated as key genes involved in carbon, nitrogen or sulfur cycling from any of the gene annotation approaches (i.e. prokka, kofam or PFAM) were considered in further analysis. The taxonomic classes were assigned to bins using the Genome Taxonomy Database classifier tool (Chaumeil *et al.*, 2020). Pathway analysis was done manually for pathways relevant to carbon fixation and nitrogen and sulfur redox reactions by integrating annotation information present in the MetaCyc (Caspi *et al.*, 2018) and KEGG databases (Kanehisa *et al.*, 2017). Completeness and contamination scores were evaluated using CheckM (Parks *et al.*, 2015), using cut-offs of  $\geq 80\%$  complete and  $\leq 10\%$  contamination. Statistics about MAGs, including classification, completeness and contamination, and MAG gene content associated with key genes in carbon, sulfur and nitrogen cycling can be found in Supporting Information Files S4 and S5.

#### Metatranscriptomic read mapping to MAGs

Relative expression of pathway-specific genes was determined by mapping reads from the metatranscriptomic data set generated by Hewson and colleagues (2014) to genes within MAGs. Reads mapping to key genes of interest within MAGs was quantified using salmon (Patro *et al.*, 2017). The average expression per sample of a variety of the following housekeeping genes was used to normalize the expression of pathway-specific genes within each MAG and sample (Nielsen and Boye, 2005; Gomes *et al.*, 2018): pyrroline-5-carboxylate reductase (*proC*), DNA recombination/repair protein (*recA*), sigma 70/D factor (*rpoD*), rho family GTPase protein (*rho*), serine hydroxymethyltransferase (*glyA*), triosephosphate isomerase (*tpiA*), DNA repair protein (*recF*). Relative fold-change in expression was calculated using the ratio of functional gene abundance to the median of all housekeeping genes detected per MAG per sample. All scripts used to process the data, as well as those used to generate the figures are posted to this project's GitHub repository (<https://github.com/spacocho/ChesapeakeSulfurOxidation>). Additional data files, including MAG and ASV sequence files, MAG functional gene annotations and expression, and ASV table, have been archived in the Johns Hopkins University data archive (<https://doi.org/10.7281/T1/NL3HTW>).

#### Phylogenetic comparison of MAG sulfur and nitrogen cycling genes

A subset of key sulfur and nitrogen cycling genes from four MAGs (CB33.bin.202, CB33.bin.222, CB33.bin.236 and CBres.bin.395) and genomes of three strains with the ability to oxidize reduced sulfur species under denitrifying conditions [*Varunaivibrio sulfuroxidans*, ASM434172v1 (Patwardhan and Vetriani, 2016); *M. blakemorei*, ASM174675v1 (Bazyliński *et al.*, 2013); *Thioalbus denitrificans*, ASM333773v1 (Park *et al.*, 2011)] were compared using phylogenetic analysis. Two representative genes involved in the oxidation of reduced sulfur species (*dsrA*, *soxX*) and two involved in the reduction of nitrogen (*nap*, *nirS*) were chosen for analysis. Genes within known sulfide-oxidizing genomes were identified with prodigal (Hyatt *et al.*, 2010) and gene annotation was done with GhostKOALA (Kanehisa *et al.*, 2016) against the KEGG database (Kanehisa *et al.*, 2017) using default settings. Genes most similar to each MAG gene in the KEGG GENE database were identified with BLASTP using an online tool (<https://www.genome.jp/tools/blast/>) and the gene sequences from the top 50 hits were downloaded. MAGs, known sulfide-oxidizing strain genomes, and KEGG genes were aligned with mafft (Kato and Standley, 2013) and a maximum likelihood tree was made with fasttree (Price *et al.*, 2010) using default settings. Trees or subtrees were visualized with NJPlot (Perriere and Gouy, 1996).

#### Statistical analysis

Principal component analysis was performed on water chemistry data using FactoMineR R package, using default settings (Husson *et al.*, 2017). To compare the variation observed in water chemistry in conjunction with microbial community variation, non-metric multi-dimensional scaling (NMDS) was performed on the Euclidean sample distances on rarefied ASV abundances after phylogenetic isometric log ratios (PhILR) transformation (Silverman *et al.*, 2017). Transformation using the PhILR was employed to account for both the compositional nature of the data and relationships between members of the population. Clustering of microbial abundances was applied to the three-dimensional decomposition of the data by NMDS. Each feature of the water quality data was centred and scaled to unit variance prior to clustering. Clustering of both data types was performed using the k-medoids algorithm on Euclidean distances (Schubert and Rousseeuw, 2019). The number of clusters within a range of two to 25 clusters was selected by calculating Hartigan's metric (Hartigan, 1975). The optimum was three for both water chemistry data and microbial compositions. Lineages

with differential abundance in space (Bay position or water column position) and time (seasons) were determined with ANCOM (Mandal *et al.*, 2015) for each taxonomic rank using QIIME2 (Bolyen *et al.*, 2019). Significance values on differentially distributed taxonomic groups were computed with Wilcoxon rank sum test (Hollander, 1999) with continuity correction in R (R Core Team, 2015).

#### Random Forest prediction of hypoxia with amplicon-based predictors

Hypoxia was encoded as a binary categorical variable derived using  $2.0 \text{ mg l}^{-1}$  DO as a cut-off. ASV counts were transformed using the centred-log-ratio transform, after replacing zeros using the *cmultRepl* function of the zCompositions R package (Xia *et al.*, 2018). These counts were aggregated by taxonomic family and FAPROTAX function. Redundant and highly correlated ( $> 0.95$ ) FAPROTAX functions were reduced to a single representative (e.g. nitrite denitrification, nitrous oxide denitrification, nitrate denitrification and denitrification were reduced to just denitrification). Random-forest based variable importance VI values were calculated using the recursive-feature elimination (*rfe*) function in the *caret* R package, since it uses cross-validation to normalize VI scores (Kuhn, 2008). The cross-validation method selected was 10 repeats on 10 folds.

#### Acknowledgements

The authors would like to acknowledge Dr. Ian Hewson for providing counsel, additional data and metadata for CB transcriptomes. SPP and KAW received support from Maryland Sea Grant under award 131160 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. Maryland Department of Natural Resources and Old Dominion Water Quality Laboratory generously provided water samples collected during their Chesapeake Bay water quality monitoring program. We thank the captain, Ed Darwin, and first mate, Jim Stickney, of the *Becky D* for their assistance with sampling at CB3.3C station. Part of this project was conducted using computational resources at the Maryland Advanced Research Computing Center.

#### References

Altieri, A.H., and Gedan, K.B. (2015) Climate change and dead zones. *Glob Chang Biol* **21**: 1395–1406.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020) KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**: 2251–2252.

Arora-Williams, K., Olesen, S.W., Scandella, B.P., Delwiche, K., Spencer, S.J., Myers, E.M., *et al.* (2018)

Dynamics of microbial populations mediating biogeochemical cycling in a freshwater lake. *Microbiome* **6**: 165.

Bazylnski, D.A., Williams, T.J., Lefevre, C.T., Trubitsyn, D., Fang, J.S., Beveridge, T.J., *et al.* (2013) *Magnetovibrio blakemorei* gen. nov., sp. nov., a magnetotactic bacterium (Alphaproteobacteria: Rhodospirillaceae) isolated from a salt marsh. *Int J Syst Evol Microbiol* **63**: 1824–1833.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N., Abnet, C.C., Al-Ghalith, G.A., *et al.* (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**: 852–857.

Bonk, F., Popp, D., Harms, H., and Centler, F. (2018) PCR-based quantification of taxa-specific abundances in microbial communities: quantifying and avoiding common pitfalls. *J Microbiol Methods* **153**: 139–147.

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., *et al.* (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**: 725–731.

Bowman, J.S., and Ducklow, H.W. (2015) Microbial communities can be described by metabolic structure: a general framework and application to a seasonally variable, depth-stratified microbial community from the Coastal West Antarctic peninsula. *PLoS One* **10**: 18.

Breitburg, D., Levin, L.A., Oschlies, A., Gregoire, M., Chavez, F.P., Conley, D.J., *et al.* (2018) Declining oxygen in the global ocean and coastal waters. *Science* **359**: 46.

Broman, E., Sachpazidou, V., Pinhassi, J., and Dopson, M. (2017) Oxygenation of hypoxic Coastal Baltic Sea sediments impacts on chemistry, microbial community composition, and metabolism. *Front Microbiol* **8**: 2453.

Bushnell, B. (2014). *BBMap: a fast, accurate, splice-aware aligner*. URL <https://www.osti.gov/servlets/purl/1241166>

Cai, W.J., Huang, W.J., Luther, G.W., Pierrot, D., Li, M., Testa, J., *et al.* (2017) Redox reactions and weak buffering capacity lead to acidification in the Chesapeake Bay. *Nat Commun* **8**: 12.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**: 581–583.

Callbeck, C.M., Lavik, G., Ferdelman, T.G., Fuchs, B., Gruber-Vodicka, H.R., Hach, P.F., *et al.* (2018) Oxygen minimum zone cryptic sulfur cycling sustained by offshore transport of key sulfur oxidizing bacteria. *Nat Commun* **9**: 11.

Canfield, D.E., Erik, K., and Bo, T. (2005b) The sulfur cycle. In *Aquatic Geomicrobiology*. Donald E. Canfield, E.K., and Bo, T. (eds). Advances in Marine Biology. Amsterdam, London: Academic Press. Advances in Marine Biology. Vol. **48**, pp. 313–381.

Canfield, D.E., Kristensen, E., and Thamdrup, B. (2005a) *Aquatic Geomicrobiology*. Advances in Marine Biology, Vol. **48**. Amsterdam, London: Academic Press.

Caspi, R., Billington, R., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., *et al.* (2018) The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res* **46**: D633–D639.

Cerco, C.F., and Noel, M.R. (2017) . In *The 2017 Chesapeake Bay Water Quality and Sediment Transport Model*,

- US Army Engineer Research and Development Center (ed). Vicksburg, MS: US Environmental Protection Agency Chesapeake Bay Program.
- Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D. H. (2020) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**: 1925–1927.
- Colatrisano, D., Ramachandran, A., Yergeau, E., Maranger, R., Gelinis, Y., and Walsh, D.A. (2015) Metaproteomics of aquatic microbial communities in a deep and stratified estuary. *Proteomics* **15**: 3566–3579.
- Crump, B.C., Peranteau, C., Beckingham, B., and Cornwell, J.C. (2007) Respiratory succession and community succession of bacterioplankton in seasonally anoxic estuarine waters. *Appl Environ Microbiol* **73**: 6802–6810.
- Dahl, C., and Prange, A. (2006) Bacterial sulfur globules: occurrence, structure and metabolism. In *Inclusions in Prokaryotes*, Shively, J.M. (ed). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 21–51.
- Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput Biol* **7**: 16.
- Eggleston, E.M., Lee, D.Y., Owens, M.S., Cornwell, J.C., Crump, B.C., and Hewson, I. (2015) Key respiratory genes elucidate bacterial community respiration in a seasonally anoxic estuary. *Environ Microbiol* **17**: 2306–2318.
- Fennel, K., and Testa, J.M. (2019) Biogeochemical controls on coastal hypoxia. *Ann Rev Mar Sci* **11**: 105–130. <https://doi.org/10.1146/annurev-marine-010318-095138>
- Findlay, A.J., Bennett, A.J., Hanson, T.E., and Luther, G.W. (2015) Light-dependent sulfide oxidation in the anoxic zone of the Chesapeake Bay can be explained by small populations of phototrophic bacteria. *Appl Environ Microbiol* **81**: 7560–7569.
- Findlay, A.J., Di Toro, D.M., and Luther, G.W. (2017) A model of phototrophic sulfide oxidation in a stratified estuary. *Limnol Oceanogr* **62**: 1853–1867.
- Flood, B.E., Jones, D.S., and Bailey, J.V. (2015) *Sedimenticola thiotaurini* sp nov., a sulfur-oxidizing bacterium isolated from salt marsh sediments, and emended descriptions of the genus *Sedimenticola* and *Sedimenticola selenitireducens*. *Int J Syst Evol Microbiol* **65**: 2522–2530.
- Friedrichs, M. (2019). 2019 Chesapeake Bay Dead Zone report; November 2019. URL [https://www.vims.edu/newsandevents/\\_docs/vims\\_chesapeakebay\\_deadzone\\_report\\_2019.pdf](https://www.vims.edu/newsandevents/_docs/vims_chesapeakebay_deadzone_report_2019.pdf)
- Ghosh, W., and Dam, B. (2009) Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. *FEMS Microbiol Rev* **33**: 999–1043.
- Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrueck, L., Reeder, J., Temperton, B., et al. (2012) Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298–308.
- Glaubitz, S., Kiesslich, K., Meeske, C., Labrenz, M., and Jurgens, K. (2013) SUP05 dominates the Gammaproteobacterial sulfur oxidizer assemblages in pelagic Redoxclines of the Central Baltic and Black Seas. *Appl Environ Microbiol* **79**: 2767–2776.
- Gomes, A.E.I., Stuchi, L.P., Siqueira, N.M.G., Henrique, J. B., Vicentini, R., Ribeiro, M.L., et al. (2018) Selection and validation of reference genes for gene expression studies in *Klebsiella pneumoniae* using reverse transcription quantitative real-time PCR. *Sci Rep* **8**: 14.
- Han, Y., Zhang, M., Chen, X.F., Zhai, W.D., Tan, E.H., and Tang, K. (2022) Transcriptomic evidences for microbial carbon and nitrogen cycles in the deoxygenated seawaters of Bohai Sea. *Environ Int* **158**: 106889.
- Hanson, T.E., Luther, G.W., Findlay, A.J., MacDonald, D.J., and Hess, D. (2013) Phototrophic sulfide oxidation: environmental insights and a method for kinetic analysis. *Front Microbiol* **4**: 8.
- Hartigan, J.A. (1975) *Clustering Algorithms*. New York: Wiley.
- He, P.Q., Xie, L.P., Zhang, X.L., Li, J., Lin, X.Z., Pu, X.M., et al. (2020) Microbial diversity and metabolic potential in the stratified Sansha Yongle Blue Hole in the South China Sea. *Sci Rep* **10**: 5949.
- Hewson, I., Eggleston, E.M., Doherty, M., Lee, D.Y., Owens, M., Shapleigh, J.P., et al. (2014) Metatranscriptomic analyses of plankton communities inhabiting surface and subpycnocline waters of the Chesapeake Bay during oxic-anoxic-oxic transitions. *Appl Environ Microbiol* **80**: 328–338.
- Hollander, M. (1999). In *Nonparametric Statistical Methods*, Wolfe, D.A. (ed). New York: Wiley.
- Hu, C.M., Li, D.Q., Chen, C.S., Ge, J.Z., Muller-Karger, F.E., Liu, J.P., et al. (2010) On the recurrent *Ulva prolifera* blooms in the Yellow Sea and East China Sea. *J Geophys Res Oceans* **115**: 8.
- Hugerth, L.W., Larsson, J., Alneberg, J., Lindh, M.V., Legrand, C., Pinhassi, J., and Andersson, A.F. (2015) Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol* **16**: 18.
- Husson, F., Le, S., and Pages, J. (2017) *Exploratory Multivariate Analysis by Example Using R*. New York: Chapman and Hall/CRC.
- Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
- Ji, B.Y., Zhang, S.D., Arnoux, P., Rouy, Z., Alberto, F., Philippe, N., et al. (2014) Comparative genomic analysis provides insights into the evolution and niche adaptation of marine *Magnetospira* sp QH-2 strain. *Environ Microbiol* **16**: 525–544.
- Kan, J.J., Crump, B.C., Wang, K., and Chen, F. (2006) Bacterioplankton community in Chesapeake Bay: predictable or random assemblages. *Limnol Oceanogr* **51**: 2157–2169.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**: D353–D361.
- Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**: 726–731.
- Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.



- Kemp, W.M., Boynton, W.R., Adolf, J.E., Boesch, D.F., Boicourt, W.C., Brush, G., *et al.* (2005) Eutrophication of Chesapeake Bay: historical trends and ecological interactions. *Mar Ecol Prog Ser* **303**: 1–29.
- Kuhn, M. (2008) Building predictive models in R using the caret package. *J Stat Softw* **28**: 1–26.
- Laas, P., Satova, E., Lips, I., Lips, U., Simm, J., Kisand, V., and Metsis, M. (2016) Near-bottom hypoxia impacts dynamics of Bacterioplankton assemblage throughout water column of the Gulf of Finland (Baltic Sea). *PLoS One* **11**: e0156147.
- Lavik, G., Stuhmann, T., Bruchert, V., Van der Plas, A., Mohrholz, V., Lam, P., *et al.* (2009) Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* **457**: 581–U586.
- Li, D.H., Liu, C.M., Luo, R.B., Sadakane, K., and Lam, T.W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674–1676.
- Li, W.L., Dong, X.Y., Lu, R., Zhou, Y.L., Zheng, P.F., Feng, D., and Wang, Y. (2021) Microbial ecology of sulfur cycling near the sulfate-methane transition of deep-sea cold seep sediments. *Environ Microbiol* **23**: 6844–6858.
- Lotze, H.K., Lenihan, H.S., Bourque, B.J., Bradbury, R.H., Cooke, R.G., Kay, M.C., *et al.* (2006) Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science* **312**: 1806–1809.
- Louca, S., Hawley, A.K., Katsev, S., Torres-Beltran, M., Bhatia, M.P., Kheirandish, S., *et al.* (2016b) Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. *Proc Natl Acad Sci USA* **113**: E5925–E5933.
- Louca, S., Parfrey, L.W., and Doebeli, M. (2016a) Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**: 1272–1277.
- Luther, G.W., Ferdelman, T., and Tsamakis, E. (1988) Evidence suggesting anaerobic oxidation of the bisulfide ion in Chesapeake Bay. *Estuaries* **11**: 281–285.
- Mandal, S., Van Treuren, W., White, R.A., Eggesbø, M., Knight, R., and Peddada, S.D. (2015) Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease* **26**: 27663.
- Mao, M.H., and Xia, M. (2020) Particle dynamics in the near-shore of Lake Michigan revealed by an observation-modeling system. *J Geophys Res Oceans* **125**: 26.
- Muller, A.L., Kjeldsen, K.U., Rattei, T., Pester, M., and Loy, A. (2015) Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi) sulfite reductases. *ISME J* **9**: 1152–1165.
- Nielsen, K.K., and Boye, M. (2005) Real-time quantitative reverse transcription-PCR analysis of expression stability of *Actinobacillus pleuropneumoniae* housekeeping genes during in vitro growth under iron-depleted conditions. *Appl Environ Microbiol* **71**: 2949–2954.
- Park, S.J., Pham, V.H., Jung, M.Y., Kim, S.J., Kim, J.G., Roh, D.H., and Rhee, S.K. (2011) *Thioalbus denitrificans* gen. nov., sp. nov., a chemolithoautotrophic sulfur-oxidizing gammaproteobacterium, isolated from marine sediment. *Int J Syst Evol Microbiol* **61**: 2045–2051.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**: 417.
- Patwardhan, S., and Vetriani, C. (2016) *Varunaivbrio sulfuroxidans* gen. nov., sp. nov., a facultatively chemolithoautotrophic, mesophilic alphaproteobacterium from a shallow-water gas vent at Tor Caldara, Tyrrhenian Sea. *Int J Syst Evol Microbiol* **66**: 3579–3584.
- Perriere, G., and Gouy, M. (1996) WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* **78**: 364–369.
- Preheim, S.P., Olesen, S.W., Spencer, S.J., Materna, A., Varadharajan, C., Blackburn, M., *et al.* (2016) Surveys, simulation and single-cell assays relate function and phylogeny in a lake ecosystem. *Nat Microbiol* **1**: 16130.
- Preheim, S.P., Perrotta, A.R., Martin-Platero, A.M., Gupta, A., and Alm, E.J. (2013) Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Appl Environ Microbiol* **79**: 6593–6603.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.
- R Core Team. (2015) *R: A Language and Environment for Statistical Computing*. In: R Foundation for Statistical Computing.
- Reed, D.C., Algar, C.K., Huber, J.A., and Dick, G.J. (2014) Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proc Natl Acad Sci USA* **111**: 1879–1884.
- Schubert, E., and Rousseeuw, P.J. (2019) *Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms*. Cham: Springer International Publishing, pp. 171–187.
- Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.
- Silverman, J.D., Washburne, A.D., Mukherjee, S., and David, L.A. (2017) A phylogenetic transform enhances analysis of compositional microbiota data. *Elife* **6**: 20.
- Smith, M.B., Rocha, A.M., Smillie, C.S., Olesen, S.W., Paradis, C., Wu, L.Y., *et al.* (2015) Natural bacterial communities serve as quantitative geochemical biosensors. *MBio* **6**: e00326-00315.
- Sorokin, D.Y., Tourova, T.P., Muyzer, G., and Kuenen, G.J. (2008) *Thiohalospira halophila* gen. nov., sp. nov. and *Thiohalospira alkaliphila* sp. nov., novel obligately chemolithoautotrophic, halophilic, sulfur-oxidizing gammaproteobacteria from hypersaline habitats. *Int J Syst Evol Microbiol* **58**: 1685–1692.
- Takai, K., Suzuki, M., Nakagawa, S., Miyazaki, M., Suzuki, Y., Inagaki, F., and Horikoshi, K. (2006) *Sulfurimonas paralvinellae* sp. nov., a novel mesophilic, hydrogen- and sulfur-oxidizing chemolithoautotroph within the Epsilonproteo-bacteria isolated from a deep-sea hydrothermal vent polychaete nest, reclassification of *Thiomicrospira denitrificans* as *Sulfurimonas denitrificans*

- comb. nov and emended description of the genus *Sulfurimonas*. *Int J Syst Evol Microbiol* **56**: 1725–1733.
- Treangen, T.J., Sommer, D.D., Angly, F.E., Koren, S., and Pop, M. (2011) Next generation sequence assembly with amos. *Curr Protoc Bioinformatics* **33**. <https://doi.org/10.1002/0471250953.bi1108s33>
- Tully, B.J., Graham, E.D., and Heidelberg, J.F. (2018) The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Scientific Data* **5**: 8.
- Uritskiy, G.V., DiRuggiero, J., and Taylor, J. (2018) MetaWRAP – a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**: 158.
- Vasimuddin, M., Misra, S., Li, H., and Aluru, S. (2019) Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 314–324. <https://doi.org/10.1109/IPDPS.2019.00041>
- Walsh, D.A., Zaikova, E., Howes, C.G., Song, Y.C., Wright, J.J., Tringe, S.G., et al. (2009) Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326**: 578–582.
- Wang, H.L., Zhang, C.L., Chen, F., and Kan, J.J. (2020) Spatial and temporal variations of bacterioplankton in the Chesapeake Bay: a re-examination with high-throughput sequencing analysis. *Limnol Oceanogr* **65**: 3032–3045.
- Wu, Y.W., Simmons, B.A., and Singer, S.W. (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**: 605–607.
- Xia, M., and Jiang, L. (2016) Application of an unstructured grid-based water quality model to Chesapeake Bay and its adjacent Coastal Ocean. *J Mar Sci Eng* **4**: 52.
- Xia, Y., Sun, J., and Chen, D.-G. (2018) *Statistical Analysis of Microbiome Data with R*. Singapore: Springer Singapore.
- Yabusaki, S.B., Wilkins, M.J., Fang, Y.L., Williams, K.H., Arora, B., Bargar, J., et al. (2017) Water table dynamics and biogeochemical cycling in a shallow, variably-saturated floodplain. *Environ Sci Technol* **51**: 3307–3317.
- Zhang, Q., Murphy, R.R., Tian, R., Forsyth, M.K., Trentacoste, E.M., Keisman, J., and Tango, P.J. (2018) Chesapeake Bay's water quality condition has been recovering: insights from a multimetric indicator assessment of thirty years of tidal monitoring data. *Sci Total Environ* **637**: 1617–1625.
- Zhou, Y.T., Scavia, D., and Michalak, A.M. (2014) Nutrient loading and meteorological conditions explain interannual variability of hypoxia in Chesapeake Bay. *Limnol Oceanogr* **59**: 373–384.

### Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Appendix S1:** Supporting Information.