Drivers of interannual 18S community composition in Diatoms and Dinoflagellates at C1 Monterey, CA

Dominic Kyle Ypil, Moss Landing Marine Laboratories (San Jose State University)

Mentors: Francisco Chavez, Kathleen Pitz, Kobun Truelove, and Jacoby Baker.

Summer 2025

Keywords: 18S, DNA Barcoding, Environmental DNA, Microbes

ABSTRACT

The Monterey Bay is a highly productive coastal ecosystem where upwelling inputs can shape microbial communities that contribute to biogeochemical cycling. This project examined the interannual shifts in phytoplankton using the 18S rRNA marker from 2008–2023 at station C1. The project combined environmental DNA (eDNA) with physicochemical data to identify drivers of community composition change. A novel QIIME2 pipeline revealed shifts in Phylum *Bacillariophyta* (diatoms) and Class *Dinophyceae* (dinoflagellates), particularly after the 2014–2016 warm anomaly. Diatom were dominated by *Pseudo-nitzschia*, whose decline after the warm anomaly correlated with silicate and nitrate depletions. In contrast, the calcareous dinoflagellate *Ensiculifera* increased in relative abundance despite decreases in pH, suggesting some resilience to acidification, potentially promoted by nutrient upwelling. These findings show the changes of microbial communities to perturbations and the need for expanded taxonomic resolution with other markers (COI, 12S), deeper vertical sampling, additional sites (M1, M2), and looking at agriculture driven eutrophication. This work advances methods for tracking planktonic responses to environmental change and subset key taxa of interest in the Monterey Bay.

INTRODUCTION

The Monterey Bay is a biologically rich and important region. The influence of the California current system drives upwelling in the region, creating a productive region for many marine species (Chavez et al. 2017). Marine microbes are important producers, consumers, and decomposers that mediate many global biogeochemical processes (Field et al. 1998; Falkowski et al. 2008). Often referred to as both the engines and engineers of an ecosystem, microbes are essential for driving ecological functions (Stal and Cretoiu 2022). However, knowledge of microbial community diversity in many systems remains limited because of their vast variability and functions (Bent and Forney 2008; Yeh and Fuhrman 2022). This creates challenges in distinguishing thousands of species involved in the cycling of inorganic and organic matter. With global warming and ocean acidification reshaping ecosystems, understanding the microbial diversity at the base of the food web becomes increasingly important due to their foundational role in structuring food webs and driving biogeochemical processes (Field et al. 1998; Falkowski et al. 2008). Therefore, to bridge this gap in knowledge we must learn to identify the organisms in an ecosystem and measure the impact their environment has on their diversity on appropriate spatiotemporal scales.

The Monterey Bay Time Series (MBTS) collects physical, biological, and chemical data at a near monthly interval around Monterey Bay since its inception in 1988. The purpose of the time series is to investigate the drivers of phytoplankton primary productivity, determine the biogeochemical process responsible, and how it varies on a spatiotemporal scale. Today the MBTS continues to sample three primary stations (C1, M1, and M2) for biological, physical, and chemical data used by many researchers (Figure 1). The addition of environmental DNA (eDNA) in 2008 has only expanded the MBTS dataset by providing information on the 12S, 18S, and COI communities of the region using molecular techniques.

Recent advances in molecular techniques, like DNA metabarcoding, provide an opportunity to survey biodiversity with a high degree of taxonomic specificity. DNA-based analysis has recently become a standard tool for identifying community composition from

environmental samples (Caron et al. 2012; Sunagawa et al. 2015; Burki et al. 2021). Sequencing of genetic barcodes—hypervariable regions of DNA—offers a more reliable and accurate method for evaluating communities than even traditional morphology-based identification (Burki et al. 2021). For eukaryotes, hypervariable regions of the conserved 18S rRNA gene (such as the V9 region) are commonly used to identify species (Amaral-Zettler et al. 2009; Burki et al. 2021). To obtain this information, seawater is filtered, and bulk DNA is extracted. Then, barcode-specific primers are used to amplify the target regions with PCR, which are then prepared as libraries for high-throughput DNA sequencing (Choi and Park 2020). The sequenced eDNA data undergoes preprocessing and quality control before being compared against a reference database for taxonomic identification. This information is useful in monitoring an ecosystem undergoing continues changes either seasonally or annually. We can identify which species response positively and negatively to changes within the environment. This knowledge is important to better understand trophic dynamics and microbially mediated biogeochemistry in many aquatic ecosystems.

For my project, I analyzed 18S eDNA data to assess the microbial community composition and environmental drivers at C1 on an inter annual time scale. I created a novel analysis pipeline to subset taxa from the larger 18S eDNA metadata. Additionally, I compared these taxa to physical, chemical, and biological changes through time. Because this project focuses on phytoplankton/microbial ecology, we used the highly conserved 18S rRNA regions to capture the diversity of microbes. Overall, this project improved our understanding drivers of microbial species composition in a highly productive region like Monterey Bay.

MATERIALS AND METHODS

Sample Collection:

Samples are collected monthly as part of the Monterey Bay time series initiative. A CTD cast from 0-200m is taken from C1, and 0-500m at M1 and M2 stations (Figure 1). Profiling (PCTD) and bottle (BCTD) measurements are taken at discrete depths. Post collection, BCTD samples for both nutrient and eDNA samples are collected from each Niskin bottle

and processed onboard. Environmental DNA samples are then concentrated onto acid washed swinnex filter cartridges (Cat. No. SX0002500) with a 0.22 µm Millipore filter (Cat. No. GVWP02500) using a peristaltic pump through Masterflex® tubing (Cat. No. EW-96419-25). Filtration is stopped when the rate is observed to slow or 1 liter has been filtered. Triplicates of each sample are taken, and filters are stored in DNA/RNA Shield Lysis Tubes (Cat. No. R1103) until DNA extraction. The nutrient BCTD samples are processed separately for nitrate, nitrite, ammonia, phosphate, chlorophyll, pH, and oxygen. Both PCTD and BCTD data are recorded and made available online (mbari.org/data/mbts-data/).

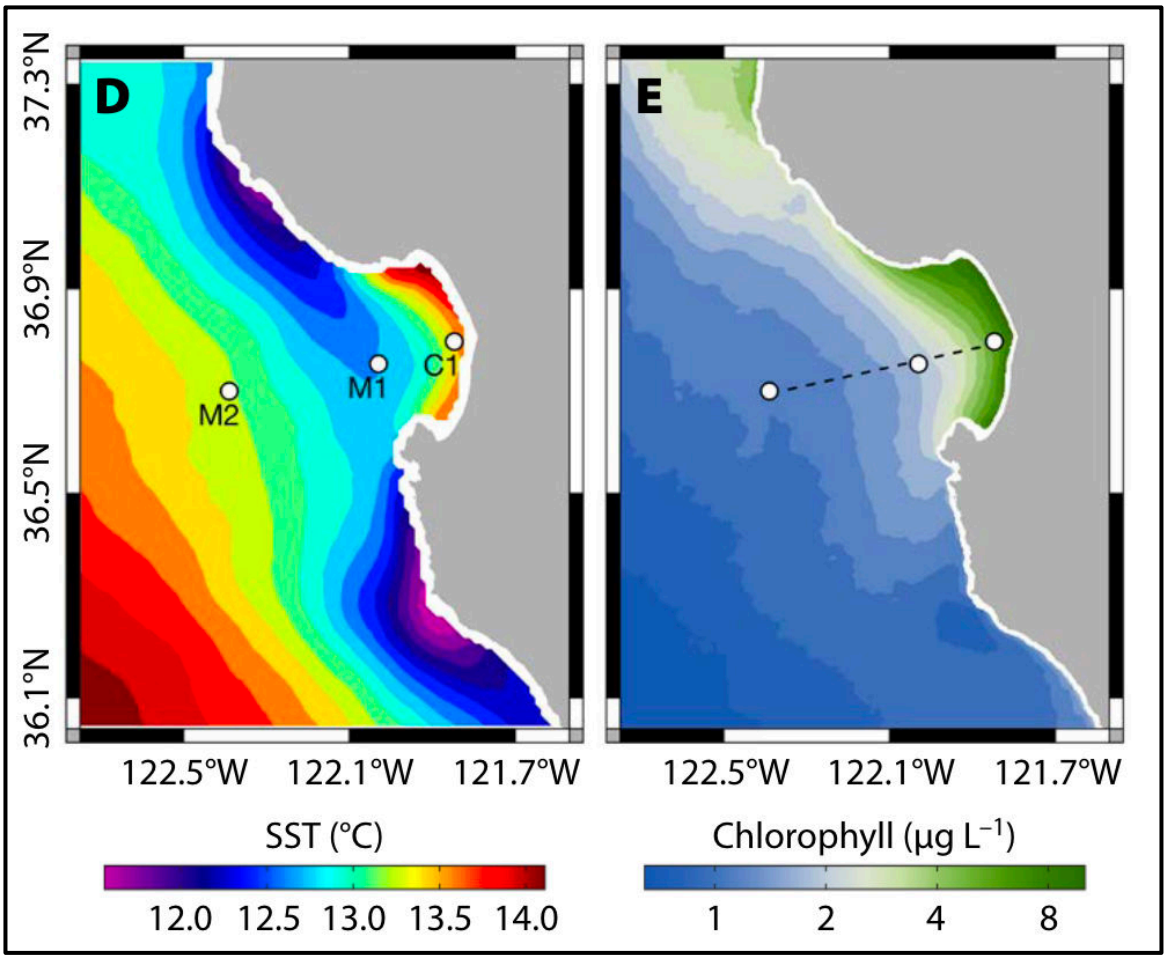

Figure 1. Average springtime D) Sea surface temperature [°C] (SST) and E) Chlorophyll concentration [µg*L^{-1}] of C1, M1, and M2 stations in Monterey Bay (Chavez et al. 2017).

Table 1: 18S ribosomal RNA primer sequences and PCR thermocycler setting.

Marker	Primers	PCR Settings	Author
18S	Forward:AATGATACGGCGACCACCGAGATCTACACTA TCGCCGTT CGGTACACACCGCCCGTC (Euk1391F) Reverse: CAAGCAGAAGACGGCATACGAGAT XXXXXXXXXXXX AGTCAGTCAG CA TGATCCTTCTGCAGGTTCACCTAC (EukBr) *(where XXXXXXXXXXXXis unique 12-bp barcode location, all primers listed in 5' to 3' direction)	10 min at 95°C For 35 cycles: * 45 sec at 94°C * 30 sec at 57°C * 90 sec at 68°C 10 min at 72°C (elongate) Hold at 4°C	(Amaral -Zettler et al. 2009)

DNA Processing and Amplicon Library Preparation:

Stored DNA samples are extracted using a Qiagen DNeasy 96 Blood and Tissue Kit (Pitz et al. 2023). Following a 96-well PCR reaction (Table 1) and purification with AMPure XP beads, the prepared amplicons are quantified (Closek et al. 2018). The samples are multiplexed (combined) and sent to a sequencing core facility at Michigan State University (East Lansing, MI, USA) for high-throughput short read sequencing on an Illumina Platform (1×150 base pairs for 18S-V9) (Guillou et al. 2013). Raw fastq files are sent back to be processed through the MBARI Banzai Pipeline (github.com/MBARI-BOG/BOG-Banzai-Dada2-Pipeline).

Bioinformatics/Data Analysis:

Once the raw fastq data is available from Michigan State University, the primer sequences are removed using atropos and placed into a bioinformatics tool called Dada2. Within Dada2, reads are trimmed for low quality, given a quality score, errors are removed, and chimeric sequences removed. The data is then assigned taxonomy through NCBI's nucleotide database and filtered using MEGAN6 set at a >80% sequence identity. From this pipeline, four files are created: 1) an amplicon sequence variant (ASV) table showing the hits of each ASV per sample, 2) a sequence table of each ASV, 3) a taxa table with the identified taxonomy of each ASV, and 4) a metadata table combining BCTD/PCTD data with sample data.

All four tables are merged in a data frame by matching samples names and ASV numbers in Python (version 3.12.11). Data is subset by station (C1), year (2008-2023), and depth (<5 meters). Graphs are generated showing percent reads over both seasonal and

inter annual time scales. Using these raw plots, taxa/groups of interest are identified for further analysis in QIIME2.

QIIME2 is bioinformatic analysis platform used for analyzing and generating plots of eDNA samples (Bolyen et al. 2019). The original metadata, ASV table, and taxa table files are converted into a BIOM file that is inputted into QIIME2 via terminal. The BIOM file is rarefied for all taxa in QIIME2 before being subset into taxa groups of interest in Python. The subset rarefied data is placed back into QIIME2 to undergo a Bray Curtis analysis to provide a distance matrix. The data is brought back to Python to calculate the principal components (PC) and plot them on a yearly timescale. A rolling average is applied to smooth out noisy seasonal fluctuations. Additionally, a biplot is created from the principal component analysis (PCA) to distinguish the top 5 ASV's driving variation. The ASVs are identified by merging a taxa table and graphed on a monthly time scale to view inter annual change.

BCTD samples are subset by station (C1), year (2008-2023), and depth (<5 meters) to be compared with taxonomy PCAs. Once the BCTD dataset is subset, parameters like nitrate, nitrite, phosphate, chlorophyll, transmission, temperature, silicate, and pH are graphed to look for trends that could explain the previous taxa PCA scores. In addition, both coastal upwelling transport index (CUTI) and biologically effective upwelling transport index (BEUTI) are plotted to account for the impact of coastal upwelling. A rolling average is applied to all parameters to smooth out any noise caused by regular seasonal fluctuation. Code for the analysis and RAW data can be found on GitHub (https://github.com/domspizza/MBARI_Internship).

RESULTS

Class Dinophyceae and Phylum Bacillariophyta:

The overall dataset had a total of 48,345,327 processed reads between 531 samples from 2008-2023. A total of 11754 unique ASVs in the data set were found.

The C1 subset (n=190) of the 18S data for *Dinophyceae* had 2741 unique ASVs while Bacillariophyta had 399. The top 20 most abundant ASVs for both *Bacillariophyta* (Figure 2) and *Dinophyceae* (Figure 3) are shown in percent reads on a monthly scale from

2008-2023. The total ASV abundance is plotted for both groups with *Bacillariophyta* (green) and *Dinophyceae* (red) with a trend line in each graph (Figure 4). The total percent reads were averaged through time to better visualize the trend and remove seasonal variability (Figure 4).

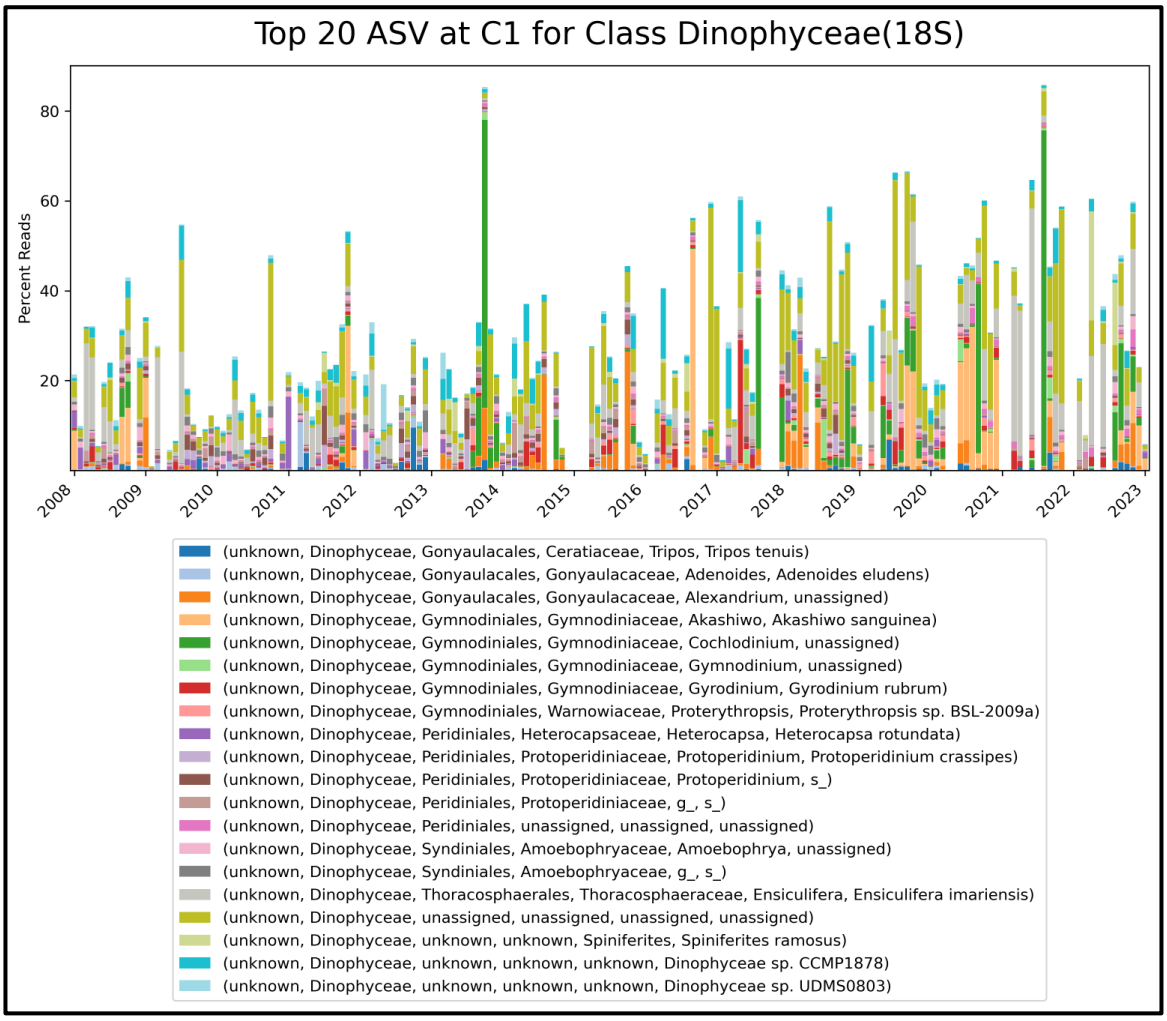

Figure 2: The top 20 ASV percent reads for the Class *Dinophyceae* from 2008-2023 at station C1.

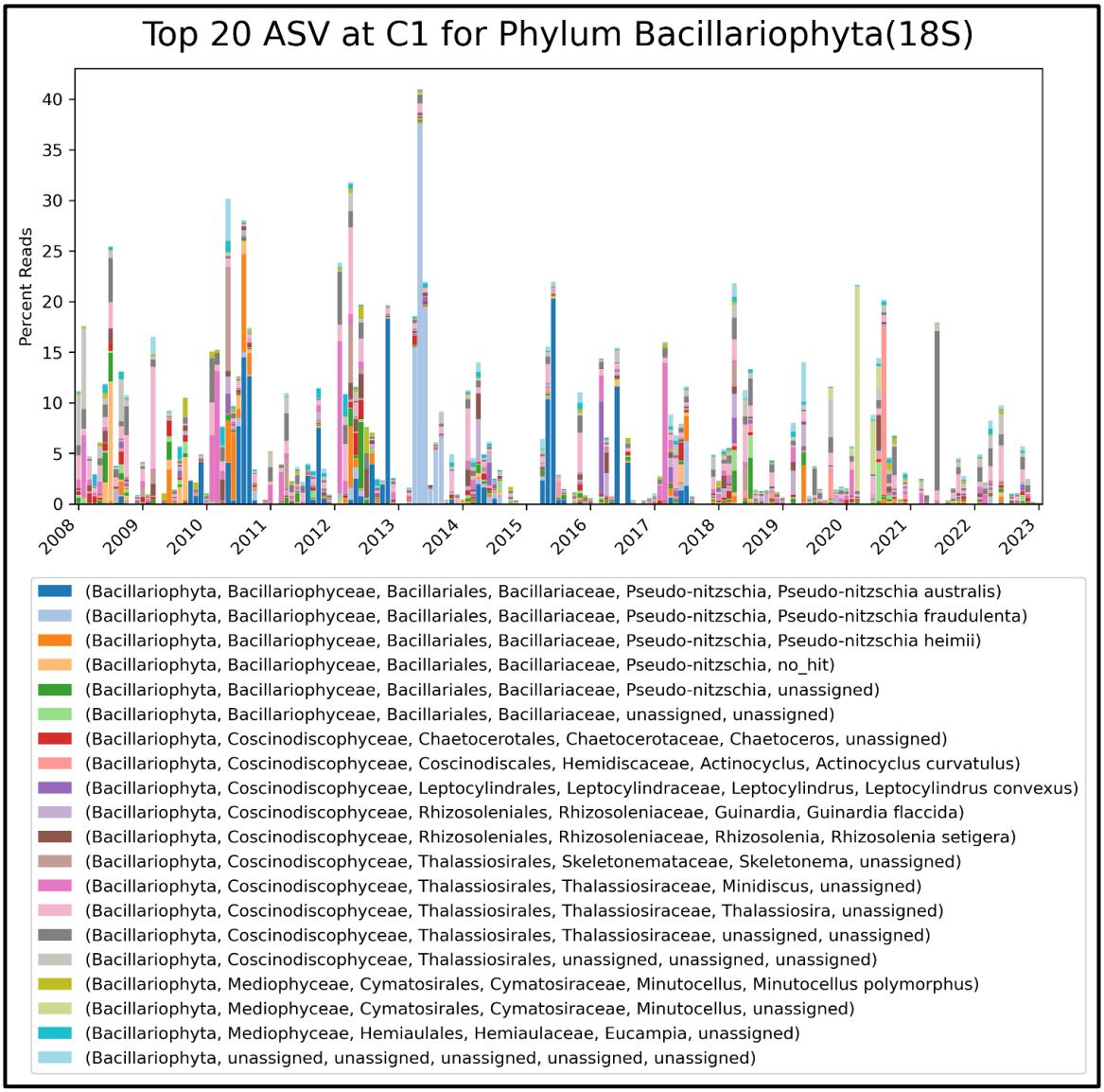

Figure 3: The top 20 ASV percent reads for the Phylum *Bacillariophyta* from 2008-2023 at station C1.

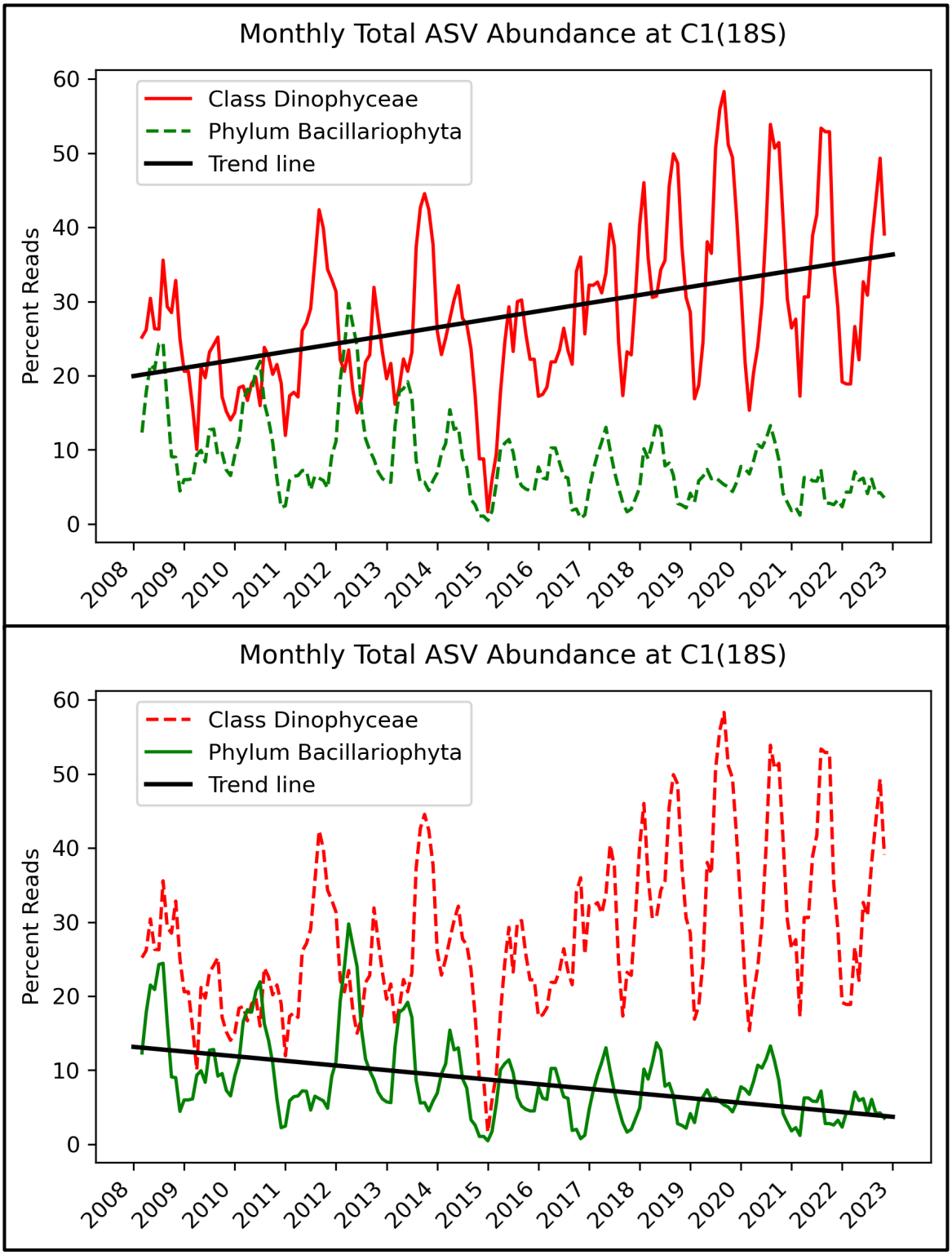

Figure 4: Total monthly percent reads for (top) Class *Dinophyceae* (red) and (bottom)Phylum *Bacillariophyta* (green) from 2008- 2023 with a trendline.

Principal Component Analysis

QIIME2 provided a distance matrix from its Bray Curtis analysis for both Class *Dinophyceae* and Phylum *Bacillariophyta*. Using SKIBIO tools, the principal component values were calculated for each group. A Principal component analysis (PCA) of Class *Dinophyceae* and Phylum *Bacillariophyta* show significant community change on the PC2 axis for both analyses (Figure 5). PC1 values were omitted as it was concluded that this was due to seasonal variation. *Bacillariophyta* had a PC1 (12.8%) and PC2 (11.2%), while *Dinophyceae* had a PC1 (9.9%) and PC2 (8.1%) (Figure 6).

PCA Biplot of both *Dinophyceae* and *Bacillariophyta* PCA show the top 5 ASVs driving variation (Figure 6). These ASVs were identified by merging with a taxa file. The Genera of *Pseudo-nitzschia* and *Ensiculifera* were found to be driving variation (Figure 7). A yearly plot of Genera for *Pseudo-nitzschia* show percent reads decreasing over time while *Ensiculifera* is increasing (Figure 7).

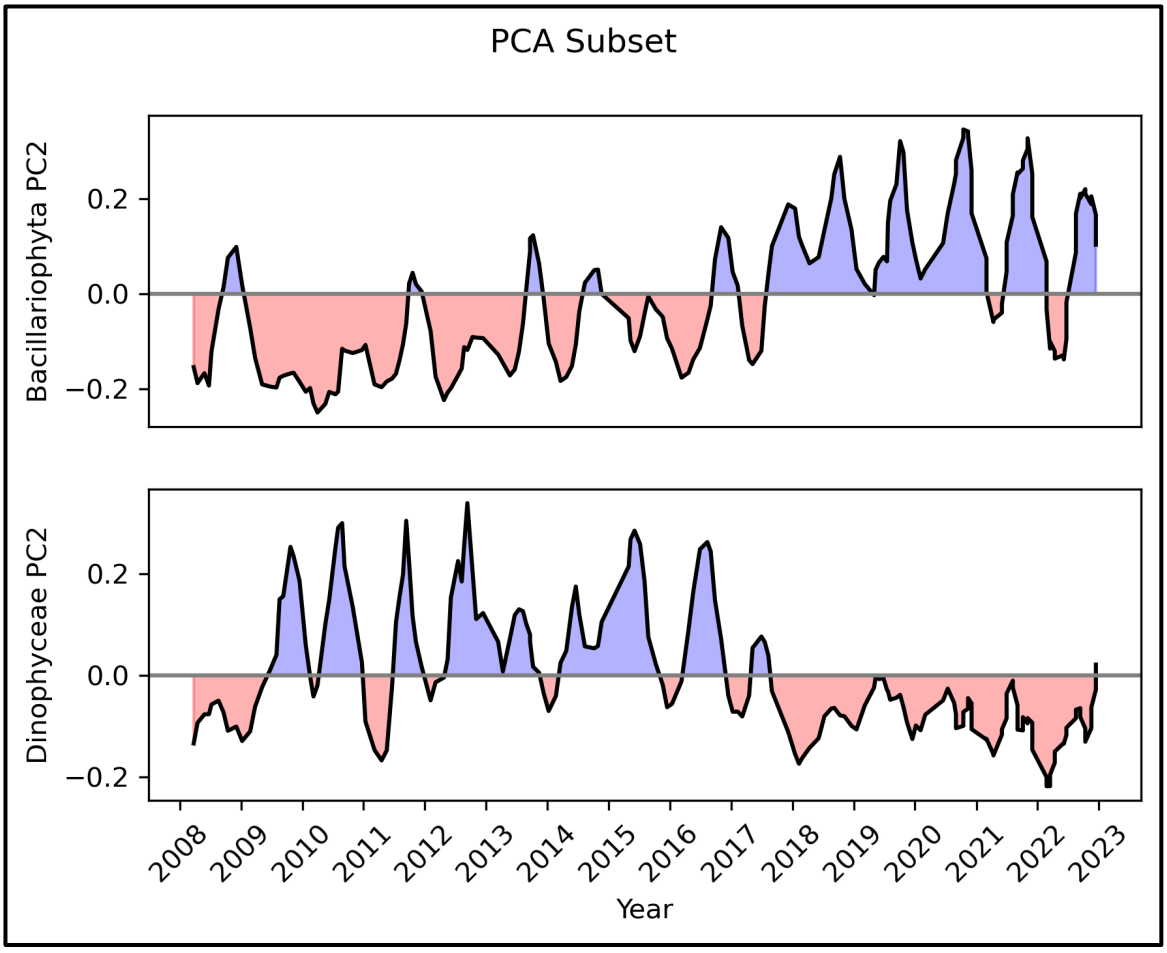

Figure 5: PC2 rolling average of both Phylum *Bacillariophyta* (top) and Class *Dinophyceae* (bottom) from 2008-2023.

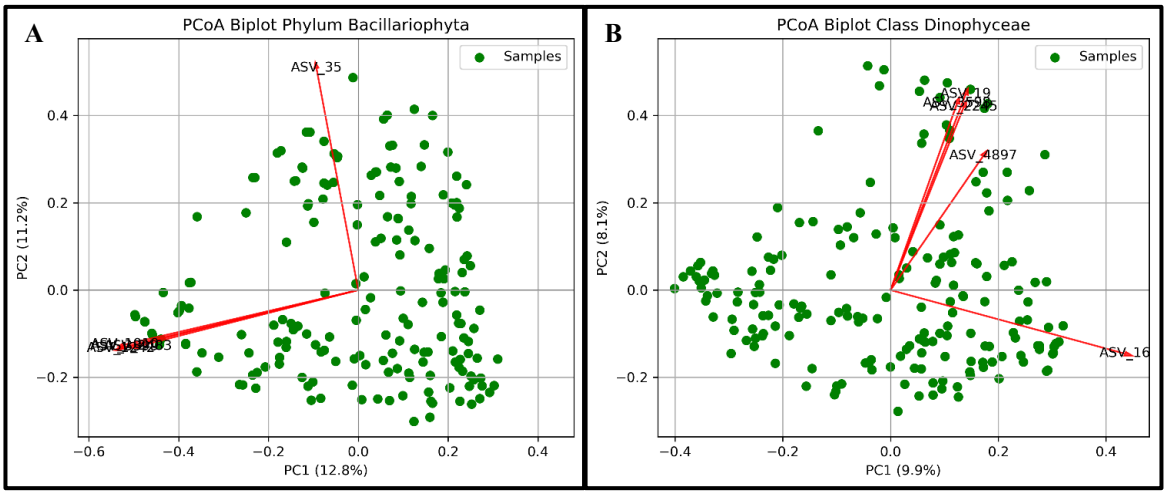

Figure 6: PCoA biplot of A) Phylum *Bacillariophyta* and B) Class *Dinophyceae* with the top 5 ASVs driving variation shown with a red arrow.

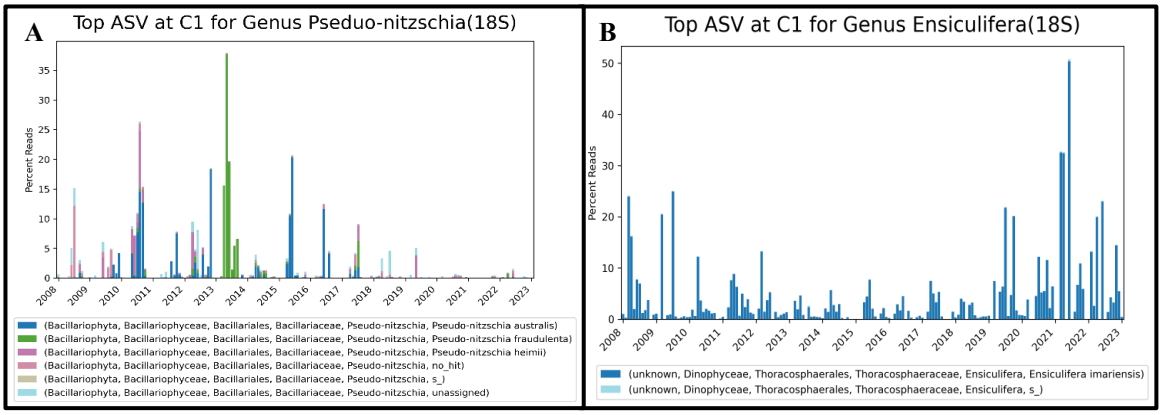

Figure 7: Monthly average percent reads of the top biplot ASVs of Genera A) *Pseudo-nitzschia* and B) *Ensiculifera* from 2008-2023.

BCTD, CUTI, and BEUITI Values

The BCTD data of key nutrients like silicate, nitrate, phosphate, and temperature are graphed from 2008-2023 for C1. The 2014-2016 warm anomaly shows a decrease in nitrate and silicate during that time. The years 2021and 2023 saw smaller cold anomaly that increased both nitrate and silicate concentrations. The anomalies were calculated by subtracting the daily average of the entire dataset with the actual values. Once the anomaly was calculated it was smoothed out by a 3-day rolling average.

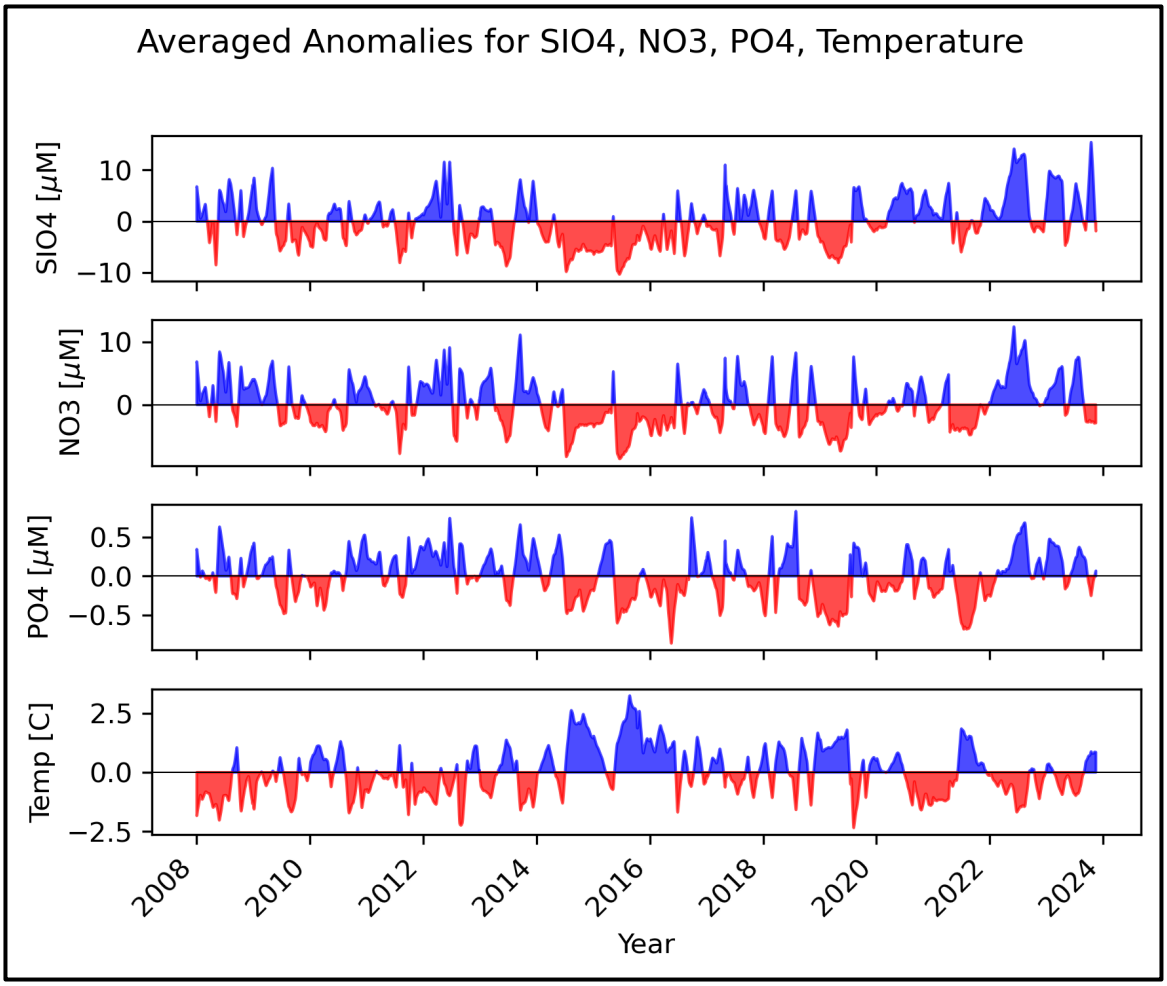

Figure 8: Bottle CTD silicate, nitrate, phosphate, and temperature values from 2008-2023 A daily average was calculated using the entire dataset and then subtracted from the actual values to remove seasonality. The anomaly was smoothed out with a rolling mean of 3 days shown in the figure.

The Coastal Upwelling Transport Index (CUTI) quantifies vertical transport near the coast, while the Biologically Effective Upwelling Transport Index (BEUTI) estimates nitrate flux (Figure 9). Both indices in the 37°N region declined during the 2014–2016 warm anomaly, reflecting reduced upwelling intensity. Following the anomaly's end, CUTI and BEUTI values required several years to rebound to pre-anomaly levels, showing a delayed recovery in upwelling-driven nutrient supply.

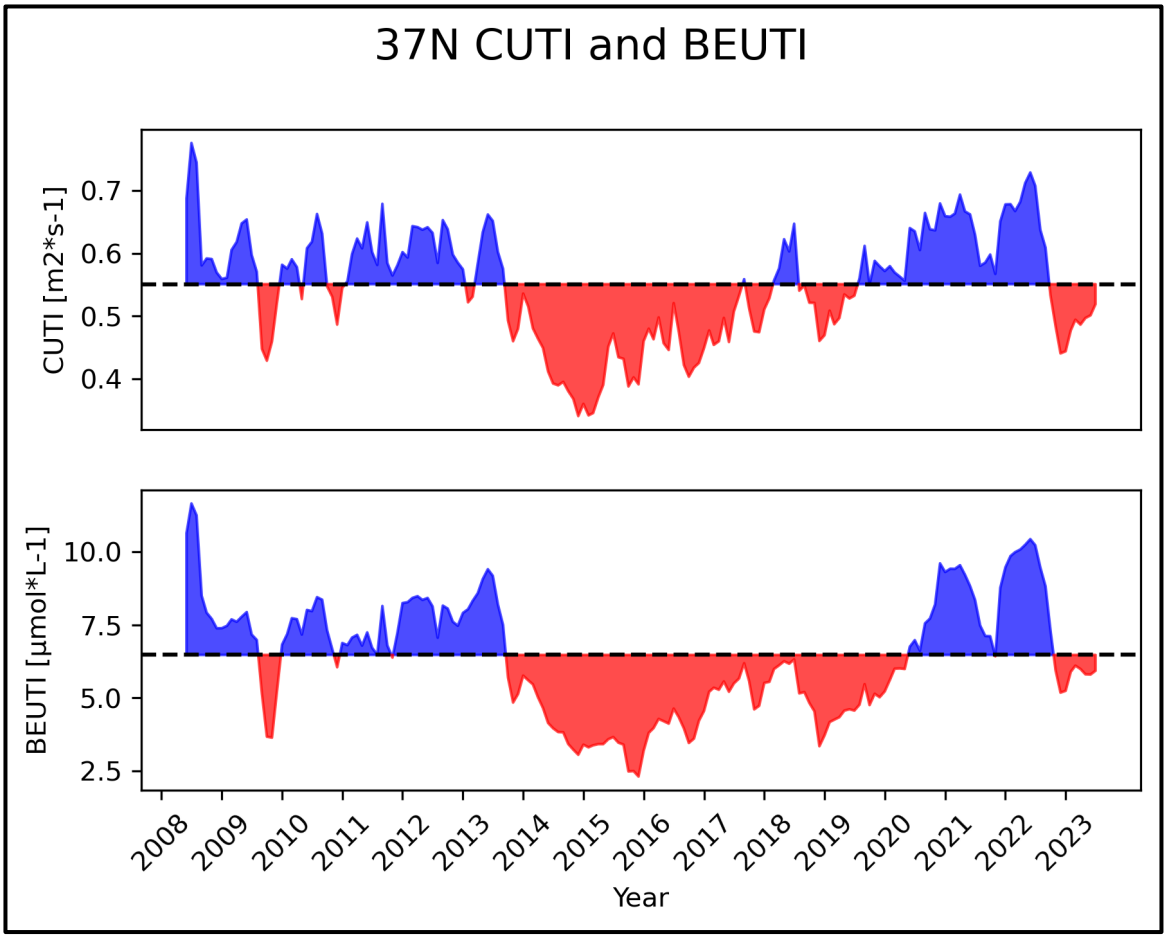

Figure 9: (Top)The Coastal Upwelling Transport Index (CUTI) and (Bottom) the Biologically Effective Upwelling Transport Index (BEUTI) from 2008-2023 at 37N. The black dashed lines represent the overall average.

DISCUSSION

This study aimed to identify the drivers of 18S community composition in Monterey Bay while developing a novel QIIME2 pipeline for analyzing eukaryotic phytoplankton taxa. PCA revealed a significant shift in the community structure of both *Bacillariophyta* (diatoms) and *Dinophyceae* (dinoflagellates), as evidenced by changes along PC2 (Figure 5). These community composition shifts were primarily driven by genera *Pseudo-nitzschia* and *Ensiculifera*. Besides *Ensiculifera,* an unknown group was also identified driving community composition, but it only accounted small portion (<5%) of the ASVs and was therefore omitted.

The genus *Pseudo-nitzschia* plays an ecologically and economically important role in Monterey Bay due to the ability of some species to produce domoic acid, a potent

neurotoxin harmful to marine life and humans. During the 2014–2016 warm anomaly, a *Pseudo-nitzschia* harmful algal bloom (HAB) occurred, lasting into 2017 (Figure 7a). Previous work has linked these blooms to silicate and nitrate stoichiometry, with silicate depletion acting as a key limiting factor during the anomaly (Ryan et al. 2017). Following the anomaly's conclusion, *Pseudo-nitzschia australis* relative abundance declined, leading to a marked shift in diatom community composition at station C1. Notably, silicate concentrations have since rebounded to pre-anomaly levels, suggesting that nutrient availability, rather than long-term depletion, was the primary driver of these dynamics. The silicate: nitrate ratio is an important factor to this as noted in Ryan et al. 2027, and could drive future HABs in the region (Ryan et al. 2017).

The calcareous dinoflagellate *Ensiculifera* exhibited an increase in relative abundance despite declining pH levels in Monterey Bay (Chavez et al. 2017). Bicarbonate is an intermediate for calcium carbonate shell formation, which is typically inhibited under low-pH conditions, the continued success of *Ensiculifera* suggests that pH was not a limited parameter. Instead, the increase in relative abundance may be linked to enhanced nutrient availability from increased vertical transport, as indicated by the rise in the Coastal Upwelling Transport Index (CUTI) and the Biologically Effective Upwelling Transport Index (BEUTI). These conditions could have brought up nutrient-rich waters to the surface, providing an ideal environment for *Ensiculifera*. The exact reason *Ensiculifera's* proliferation over other phytoplankton is still in question and would require additional information not available in this dataset.

CONCLUSIONS/RECOMMENDATIONS

From 2008 to 2023, the 18S community composition of the Class *Dinophyceae* and Phylum *Bacillariophyta* in Monterey Bay underwent a significant shift in 2017. This transition was primarily driven by changes in the relative abundance of key genera—*Pseudo-nitzschia* (diatoms) and *Ensiculifera* (dinoflagellates). The decline of *Pseudo-nitzschia* following the 2014–2016 warm anomaly suggests that the return to cooler conditions created unfavorable growth conditions for this genus. In contrast, *Ensiculifera* thrived despite decreasing pH levels, demonstrating a resilience to ocean acidification that may be linked to reduced

reliance on calcification. However, the high proportion of unassigned taxa within Dinophyceae highlights a gap in taxonomic resolution, leaving a substantial portion of this group's ecological role unaccounted for in driving community variation.

To build upon these findings, future work should incorporate additional genetic markers, such as COI and 12S, from the same samples to improve taxonomic resolution and ecological insight. Expanding sampling efforts beyond station C1 to include other key sites like M1 and M2 would provide a better understanding of community dynamics across Monterey Bay. Additionally, employing another reference databases like BOLD or PR2 may help resolve unassigned *Dinophyceae* taxa that NCBI failed to classify.

Further investigations should also explore vertical distribution patterns by extending eDNA sampling to deeper depths, as current data are largely limited to surface waters. The silicate: nitrate ratios played a large role in the formation of the HABs, looking into their relative rations could provide a better understanding of these drivers in diatoms. In addition, assessing the biological response to changes in alkalinity and carbonate chemistry of the water could show how communities shift. Finally, extending this research to adjacent ecosystems like Elkhorn Slough, a freshwater-influenced estuary experiencing agricultural eutrophication could reveal parallel or contrasting community responses to environmental stressors.

ACKNOWLEDGEMENTS

Heartfelt thanks to everyone who contributed to this incredible internship experience. It has elevated my career and confidence as a scientist to a new level. I am deeply grateful to the entire Biological Oceanography Group (BOG): Francisco Chavez, Jacoby Baker, and Marguerite Blum for their wisdom and support. A special thank you to Kathleen Pitz and Kobun Truelove for guiding my project, and welcoming me to the MBARI community. The MBARI Summer Internship Program is generously supported through a gift from the Dean and Helen Witter Family Fund and the Rentschler Family Fund in memory of former MBARI board member Frank Roberts (1920-2019) and by the David and Lucile Packard Foundation. Additional funding is provided by the Maxwell/Hanrahan Foundation.

REFERENCES

[1] Chavez, F.P., J.T. Pennington, R.P. Michisaki, M. Blum, G.M. Chavez, J. Friederich, B. Jones, R. Herlien, B. Kieft, B. Hobson, A.S. Ren, J. Ryan, J.C. Sevadjian, C.

Wahl, K.R. Walz, K. Yamahara, G.E. Friederich, and M. Messié. 2017. Climate variability and change: Response of a coastal ocean ecosystem. Oceanography 30(4):128–145. doi.org/10.5670/oceanog.2017.429.

[2] Amaral-Zettler, L. A., E. A. McCliment, H. W. Ducklow, and S. M. Huse. 2009. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. PLOS ONE **4**: e6372. doi:10.1371/journal.pone.0006372

[3] Bent, S. J., and L. J. Forney. 2008. The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. ISME J **2**: 689–695. doi:10.1038/ismej.2008.44

[4] Bolyen, E. and others. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol **37**: 852–857. doi:10.1038/s41587-019-0209-9

[5] Burki, F., M. M. Sandin, and M. Jamy. 2021. Diversity and ecology of protists revealed by metabarcoding. Current Biology **31**: R1267–R1280. doi:10.1016/j.cub.2021.07.066

[6] Caron, D. A., P. D. Countway, A. C. Jones, D. Y. Kim, and A. Schnetzer. 2012. Marine Protistan Diversity. Annu. Rev. Mar. Sci. **4**: 467–493. doi:10.1146/annurev-marine-120709-142802

[7] Choi, J., and J. S. Park. 2020. Comparative analyses of the V4 and V9 regions of 18S rDNA for the extant eukaryotic community using the Illumina platform. Sci Rep **10**: 6519. doi:10.1038/s41598-020-63561-z

[8] Closek, C. and others. 2018. Environmental DNA (eDNA) 18S metabarcoding Illumina MiSeq NGS PCR Protocol.

[9] Falkowski, P. G., T. Fenchel, and E. F. Delong. 2008. The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science **320**: 1034–1039. doi:10.1126/science.1153213

[10] Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science **281**: 237–240. doi:10.1126/science.281.5374.237

[11] Guillou, L. and others. 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res **41**: D597–D604. doi:10.1093/nar/gks1160

[12] Pitz, K., J. Baker, and truelove. 2023. Environmental DNA (eDNA) extraction using Qiagen DNeasy 96 Blood and Tissue Kit.

[13] Ryan, J. P. and others. 2017. Causality of an extreme harmful algal bloom in Monterey Bay, California, during the 2014–2016 northeast Pacific warm anomaly. Geophysical Research Letters **44**: 5571–5579. doi:10.1002/2017GL072637

[14] Stal, L. J., and M. S. Cretoiu, eds. 2022. The Marine Microbiome, Springer International Publishing.

[15] Sunagawa, S. and others. 2015. Structure and function of the global ocean microbiome. Science **348**: 1261359. doi:10.1126/science.1261359

[16] Yeh, Y.-C., and J. A. Fuhrman. 2022. Contrasting diversity patterns of prokaryotes and protists over time and depth at the San-Pedro Ocean Time series. ISME COMMUN. **2**: 1–12. doi:10.1038/s43705-022-00121-8