UNIVERSITY OF CALIFORNIA

Santa Barbara

Phytoplankton community determinations and dynamics in the Santa Barbara Channel, California

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Marine Science

by

Dylan Samuel Catlett

Committee in charge:

Professor David A. Siegel, Chair

Professor M. Debora Iglesias-Rodriguez

Professor Elizabeth G. Wilbanks

September 2021

The dissertation of Dylan Samuel Catlett is approved.

_____

Professor Elizabeth G. Wilbanks

_____

Professor M. Debora Iglesias-Rodriguez

_____

Professor David A. Siegel, Committee Chair

August 2021

Phytoplankton community determinations and dynamics in the Santa Barbara Channel, California

Dylan Samuel Catlett

ACKNOWLEDGEMENTS

This work was funded predominantly by a NASA Earth and Space Science Fellowship and the Santa Barbara Channel Marine Biodiversity Observation Network (jointly funded by NASA, NOAA, and BOEM). Most of the data presented here were collected as part of the NASA-funded Plumes and Blooms project, with ship time supported by the Channel Islands National Marine Sanctuary. Additional financial support was provided by the NSF-funded Santa Barbara Coastal Long Term Ecological Research program, the UCSB Coastal Fund, and the UCSB Earth Research Institute.

I am immensely grateful for the scientific and professional support I received from many folks during my time at UCSB. Special thanks to the Plumes and Blooms team, especially Nathalie Guillocheau, Stuart Halewood, and Emma Wear, for collecting most of the samples and data this dissertation was built on. Thanks to my lab-mates in both the Siegel and Iglesias-Rodriguez groups for many helpful conversations, sharing data and code, and generally teaching me how to do science. More IGPMS faculty, researchers, and students than I can name here were tremendously helpful and supportive of this work and my professional development as an oceanographer. Thanks especially to Lizzy Wilbanks and Craig Carlson for regularly providing excellent science and career advice, and to Bob Miller for giving me loads of opportunities through the MBON. Finally, thank you to Dave Siegel and Debora Iglesias-Rodriguez for giving me the opportunity to come to UCSB in the first place, and for the constant support and guidance since I arrived. I really couldn't imagine having more supportive mentors.

Most important, this dissertation is dedicated to my family: Momma, Dad, Marley, Grandmas Denyse and Nancy, my late grandfathers Bucky and Ralph, and Logan and

Boogey. You all worked so hard to give me the privilege of pursuing a career in oceanography. There's no way to really say thank you for that, but I hope I've made you proud so far.

Vita of Dylan Samuel Catlett

EDUCATION

Ph.D., Marine Science, University of California, Santa Barbara, expected September 2021

B.S. Biology, B.A. Chemistry, University of North Carolina, Chapel Hill, 2014

PUBLICATIONS

* = Co-first author; [U] = undergraduate mentee

2021 **Catlett, D.**, K. Son[U], C. Liang[U]. ensembleTax: An R package for determinations of ensemble taxonomic assignments of phylogenetically-informative marker gene sequences. *PeerJ*. doi: doi.org/10.7717/peerj.11865.

2021 **Catlett, D.**, D. A. Siegel, R. D. Simons, N. Guillocheau, F. Henderikx-Freitas, C. S. Thomas. Diagnosing seasonal to multi-decadal phytoplankton group dynamics in a highly productive coastal ecosystem. *Progress in Oceanography*. doi: 10.1016/j.pocean.2021.102637.

2021 Kahru, M., C. R. Anderson, A. D. Barton, M. Carter, **D. Catlett**, U. Send, H. M. Sosik, E. L. Weiss, B. G. Mitchell. Satellite detection of dinoflagellate blooms off California by UV reflectance ratios. *Elementa: Science of the Anthropocene*. doi: 10.1525/elementa.2020.00157.

2020 Bisson, K.*, N. Baetge*, S. Kramer, **D. Catlett**, 19 additional co-authors. California wildfire burns boundaries between art and science. *Oceanography*. doi: 10.5670/oceanog.2020.110.

2019 **Catlett, D.***, P. G. Matson*, C. A. Carlson, E. G. Wilbanks, D. A. Siegel, M. D. Iglesias-Rodriguez. Evaluation of accuracy and precision in an amplicon sequencing workflow for marine protist communities. *Limnology and Oceanography: Methods*. doi: 10.1002/lom3.10343.

2018 **Catlett, D.**, and D. A. Siegel. Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment. *Journal of Geophysical Research: Oceans*. doi: 10.1002/2017JC013195.

2015 Marchetti, A., **D. Catlett**, B. M. Hopkinson, K. Ellis, N. Cassar. Marine diatom proteorhodopsins and their potential role in coping with low iron availability. *The ISME Journal.* doi: 10.1038/ismej.2015.74.

SOFTWARE AND DATA

[U] = undergraduate mentee

2021 **Catlett, D.**, K. Son[U], C. Liang[U]. ensembleTax: Ensemble Taxonomic Assignments of Amplicon Sequencing Data. R package available on Github and CRAN.

2020 **Catlett, D.**, D. A. Siegel, N. Guillocheau. Plumes and Blooms: Curated oceanographic and phytoplankton pigment observations ver 1. Environmental Data Initiative. https://doi.org/10.6073/pasta/f88ee1dc32b8785fe6ce57d80722e78c.

2019 Santa Barbara Channel 18S-V9 Illumina MiSeq raw sequence reads. National Center for Biotechnology Information Sequence Read Archive accession number PRJNA532583. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA532583.

PRESENTATIONS

[U] = undergraduate mentee

2020 **Catlett**, **D.,** D. A. Siegel, C. A. Carlson, Matson, P. G., E. K. Wear, M. D. Iglesias-Rodriguez. Integrating high-throughput sequencing observations into remotely sensible phytoplankton functional type determinations. Oral Presentation, Ocean Sciences Meeting, San Diego, CA.

2020 Matson, P. G., **D. Catlett**, C. A. Carlson, T. S. Lankiewicz, D. A. Siegel, E. K. Wear, M. D. Iglesias-Rodriguez. Seasonal Patterns in Protist Diversity Across Depths in the Santa Barbara Channel. Oral Presentation, Ocean Sciences Meeting, San Diego, CA.

2019 **Catlett**, **D.,** R. Simons, N. Guillocheau, F. Henderikx-Freitas, D. A. Siegel. Seasonal to decadal scale phytoplankton community variations in the Santa Barbara Channel, CA, are associated with decadal climate oscillations and changes in regional ocean circulation. Oral Presentation, Eastern Pacific Ocean Conference, Fallen Leaf Lake, CA.

2019 Cabrera, C.[U], **D. Catlett**, M. D. Iglesias-Rodriguez. Does wildfire ash deposition impact phytoplankton communities in the Santa Barbara Channel? Poster Presentation, UC Santa Barbara Undergraduate Research Colloquium, Santa Barbara, CA.

2018 **Catlett, D.** and D. A. Siegel. Investigating drivers of coastal phytoplankton community variations using a combination of measured and bio-optically modeled phytoplankton pigment concentrations. Poster Presentation, Ocean Sciences Meeting, Portland, OR.

2018 Bisson, K., S. Kramer, A. Fischer, **D. Catlett**, J. Allen, and D. A. Siegel. Temporal patterns and optical analysis of wildfire-derived ash in the Santa Barbara Channel. Poster Presentation, Ocean Optics XXIV, Dubrovnik, Croatia.

2017 **Catlett, D.**, D. A. Siegel, and N. Guillocheau. Derivative analysis demonstrates the potential and limitations for deriving phytoplankton community structure from hyperspectral ocean color observations. Poster Presentation, Aquatic Sciences Meeting, Honolulu, HI.

2014 **Catlett, D.** and A. Marchetti. Characterization of proteorhodopsins in the oceanic diatom *Pseudo-nitzschia granii.* Oral Presentation, UNC-Chapel Hill Biology Honors Symposium, Chapel Hill, NC.

2014 **Catlett, D.** and A. Marchetti. Characterization of proteorhodopsins in the oceanic diatom *Pseudo-nitzschia granii.* Oral Presentation, UNC-Chapel Hill Celebration of Undergraduate Research, Chapel Hill, NC.

GRANTS, AWARDS, FELLOWSHIPS, AND HONORS

2016-2019	NASA Earth and Space Science Fellowship ($125,000)
2019	UC Santa Barbara Doctoral Student Travel Grant Award ($400)
2019	UC Irvine Microbiomes and Global Change Summer Institute Scholarship
2019	UC Santa Barbara Coastal Fund Grant: "Exploring plankton diversity in the Santa Barbara Channel" ($2,751.30)
2018, 2019	Earth Research Institute Graduate Student Conference Travel Award ($500, $400)
2016	Ocean Carbon and Biogeochemistry Program Scholarship to attend Cornell University's Ocean Color Remote Sensing Training Course ($2,800)
2014	NSF Research Experience for Undergraduates ($5,000)

TEACHING AND MENTORSHIP

* = Online instruction

Instructor of Record, Department of Geography, UC Santa Barbara

Geography 3: Oceans and Atmosphere*	Winter 2021

Teaching Assistant, Department of Geography, UC Santa Barbara

Geography 165: Ocean Waves and Tides*	Spring 2020
Geography 20: Geography of Surfing	Winter 2020
Geography 165: Ocean Waves and Tides	Winter 2015
Geography 104: Physical Oceanography	Fall 2015

ACADEMIC SERVICE

Peer Reviewer

Limnology and Oceanography: Methods, one manuscript	2021
Deep Sea Research Part. I, one manuscript	2020

IGPMS Seminar Series Coordinator, UC Santa Barbara 2019-2020

IGPMS Social Committee, UC Santa Barbara 2017-2018

ABSTRACT

Phytoplankton community determinations and dynamics in the Santa Barbara Channel, California

by

Dylan Samuel Catlett

Quantifying and determining the oceanographic and climate forcings of variability in phytoplankton community composition (PCC) represents a critical step in understanding and predicting marine ecosystem structure and function. In this dissertation, I analyzed samples and data from an extensive oceanographic time series from the highly productive Santa Barbara Channel, CA (SBC). The goals of this work were to (1) determine the "best-case scenario" for retrieving PCC from remotely sensible hyperspectral bio-optical observations; (2) characterize the dominant oceanographic and climate forcings of PCC variations in the SBC; and (3) integrate HPLC pigment and amplicon sequencing observations of PCC to assess the agreement between methods and the potential to gain novel insights into phytoplankton physiology and ecology and ecosystem function.

In Chapter 2, we considered nearly a decade of concurrent HPLC phytoplankton pigment and spectrophotometric phytoplankton absorption coefficient observations. We implemented covariance-based analyses to identify 3-5 phytoplankton pigment communities from the HPLC pigment data, and spectral derivative analysis of phytoplankton absorption

to identify phytoplankton absorption features. We then developed a bio-optical modeling approach to retrieve phytoplankton pigment concentrations and PCC indices from phytoplankton absorption derivative spectra with high skill. Notably, we found that absorption features from across the visible spectrum are useful in modeling PCC.

In Chapter 3, we applied the bio-optical models from Chapter 2 to the PnB archive of phytoplankton absorption observations to double the length of the PnB HPLC phytoplankton biomarker pigment time series to 20+ years of observations. We then characterized the dominant oceanographic and climate forcings of five phytoplankton groups resolved from HPLC pigment data. Despite the widely documented dominant response of diatoms to seasonal upwelling, our analysis revealed that prymnesiophytes and chlorophytes are the typical "first responders" to seasonal upwelling in the SBC. On multi-decadal time scales, we identified an association between anomalous SBC dinoflagellate blooms and surface ocean advection patterns likely forced by a dominant mode of decadal climate variability, the North Pacific Gyre Oscillation.

Finally, Chapter 4 focused on integrating HPLC pigment and amplicon sequencing observations to assess PCC in the coastal ocean. High uncertainty in estimates of biomass contributions of specific phytoplankton groups was often introduced by inter- and intra-group variability in biomarker pigment expression. However, we observed that distinct suites of biomarker pigments covary with unique communities of phytoplankton and other protistan groups. Integrating the two data sets thus provided novel insights into phytoplankton physiology and function, and suggested a path toward monitoring ecosystem structure and function on unprecedented spatiotemporal scales via ocean color remote sensing of phytoplankton biomarker pigment concentrations.

I. Introduction

Phytoplankton account for most of the primary production in the ocean and thus play an integral role in marine food webs and biogeochemical cycles (Field et al., 1998; Guidi et al., 2016). Coastal marine ecosystems contribute disproportionately to global phytoplankton productivity despite their limited spatial extent relative to the global ocean (Ryther, 1969). The composition of the phytoplankton community in both the coastal and open ocean influences the fate of phytoplankton primary production (Ryther, 1969; Guidi et al., 2016): will it be transferred up the food web to sustain diverse and economically important marine macrofauna, or will it escape to the deep ocean and contribute to the sequestration of carbon for decades to millenia? Quantifying and determining the dominant oceanographic and climate forcings of phytoplankton community composition (PCC) in the coastal ocean on large spatiotemporal scales represents a critical step in answering this question, and in turn, improving our capacity to understand and predict marine ecosystem structure and function.

The high morphological, functional, and genetic diversity amongst marine phytoplankton makes quantifying PCC coherently across different scales of space, time, and diversity extremely difficult. Despite this, many methods with unique strengths and weaknesses are available to quantify PCC, including microscopic and/or image-based cell identification and enumeration, flow cytometry, amplicon sequencing of phylogenetically informative barcode genes, and High Performance Liquid Chromatography analysis of phytoplankton biomarker pigment concentrations (see Johnson and Martiny, 2015 for a review). PCC can also be monitored on larger spatiotemporal scales via satellite ocean color remote sensing, which relies on algorithms that take advantage of the linkages

between ocean bio-optical properties observed by ocean color satellites and in-water estimates of PCC (most often based on HPLC pigment analysis; IOCCG, 2014).

HPLC phytoplankton pigment analysis measures the concentrations of ~25 accessory pigments, some of which serve as biomarkers for specific phytoplankton groups (Van Heukelem and Thomas, 2001; Jeffrey et al., 2011). Phytoplankton pigment chemotaxonomy seeks to estimate the contributions of various phytoplankton groups to the phytoplankton community. Pigment chemotaxonomy is the most widely used method to validate ocean color and bio-optical PCC algorithms due to its rigorously validated and standardized sample analysis procedures and direct links to bio-optical properties (IOCCG, 2014). CHEMTAX and diagnostic pigment methods (Mackey et al., 1996; Vidussi et al., 2001) are the most widely used chemotaxonomic methods and provide estimates of the contributions of various phytoplankton types to the bulk chlorophyll *a* concentration. However, chemotaxonomy is associated with inherent uncertainty due to ambiguity in the representation of specific phytoplankton groups by various biomarker pigments (Jeffrey et al., 2011) and the complicated relationships between pigment concentrations and carbon biomass, primary productivity, and cell abundances that arise due to variability in phytoplankton physiological status and other factors (Goericke and Montoya 1998; Behrenfeld et al., 2005). These sources of uncertainty tend to limit the taxonomic resolution of pigment-based PCC assessments to approximately the class level (e.g., diatoms, dinoflagellates, etc.). More refined "pigment types" have been proposed (e.g., diatom type 1, diatom type 2, etc.) (Mackey et al., 1996; Higgins et al., 2011), but there is little evidence to suggest these refined pigment classes are genetically, functionally, or ecologically unique, and the ambiguity in pigment biomarkers generally

precludes accurate quantification of more than a few broadly-defined groups from HPLC pigment data.

Many approaches have been proposed to retrieve the concentrations of HPLC phytoplankton pigments and/or derived indices of PCC from ocean bio-optical properties (see Mouw et al., 2017 for a review). Spectral-based approaches for retrieving PCC indices from bio-optical properties rely on direct links between PCC indices and spectral signatures, while abundance-based approaches attempt to model PCC as a function of bio-optically modeled phytoplankton biomass (Mouw et al., 2017). The upcoming launch of NASA's Plankton, Aerosol, Cloud, and ocean Ecosystem (PACE) mission is expected to improve spectral-based retrievals of PCC from satellite ocean color estimates as it will offer improved resolution of phytoplankton optical signatures via the hyperspectral Ocean Color Instrument (Werdell et al., 2019), and thus offers a unique opportunity to observe PCC on unprecedented spatiotemporal scales. Promising spectral-based approaches for identifying and enhancing the signal arising from phytoplankton pigment absorption features in hyperspectral bio-optical observations include gaussian decomposition (Chase et al., 2017) and spectral derivative analysis (Bidigare, 1989), though further research is needed to assess the potential to quantitatively model PCC indices following isolation of phytoplankton absorption features.

DNA meta-barcoding is now widely applied throughout the world's oceans to characterize the composition of phytoplankton and other microbes (de Vargas et al., 2015). DNA meta-barcoding estimates the relative sequence abundances of amplicon sequence variants (ASVs), a proxy for microbial species (Callahan et al., 2016), as the fraction of the total sequence reads recovered from each sample that matches the ASV

sequence (Balint et al., 2016). In general, analytical uncertainty in DNA meta-barcoding analysis is poorly constrained and difficult to evaluate (Wear et al., 2018; Yeh et al., 2021). However, DNA meta-barcoding offers much finer taxonomic resolution of PCC than HPLC pigment analysis. Both DNA meta-barcoding and HPLC pigment analysis provide relatively holistic estimates of PCC across the complete range of phytoplankton size classes, and provide estimates of PCC that are more similar to one another than to PCC estimates from other methods (Gong et al., 2020). Thus, integrating HPLC pigment and DNA meta-barcoding observations is needed to ensure robust and coherent assessments of relationships between PCC and marine ecosystem structure and function.

The Santa Barbara Channel, CA (SBC) is a particularly productive, biodiverse coastal ecosystem found at the northern boundary of the Southern California Bight (Beers, 1986). The SBC lies in between the relatively cool, productive waters of the southern California Current System and the warmer, more oligotrophic waters of the Southern California Bight, and the confluence of these distinct water masses drives high spatiotemporal variability in physical, chemical, and biological oceanographic properties in the SBC (Henderikx-Freitas et al., 2017). Seasonal upwelling is the dominant oceanographic forcing of phytoplankton biomass and community dynamics in the SBC, though the persistent SBC eddy can act to further concentrate phytoplankton biomass and productivity in the SBC (Harms and Winant, 1998; Brzezinski and Washburn, 2011; Venrick, 2012; Henderikx-Freitas et al., 2017). Upwelling typically leads to substantial accumulation of diatom biomass in the SBC surface ocean, though blooms of other phytoplankton including dinoflagellates, prymnesiophytes, and chlorophytes have also been documented in or around the SBC (Countway and Caron, 2006; Venrick, 2012;

Taylor et al., 2015; Matson et al., 2019). On longer time scales, climate forcing by the major modes of North Pacific climate variability, the El Niño Southern Oscillation, Pacific Decadal Oscillation, and North Pacific Gyre Oscillation, are thought to drive interannual to decadal variability in phytoplankton biomass although the impacts of climate forcings on phytoplankton communities are under-explored (Mantua et al., 1997; Di Lorenzo et al., 2008; Venrick, 2012; Di Lorenzo et al., 2013; Jacox et al., 2016).

In this dissertation, phytoplankton community composition and dynamics in the SBC are quantified using an extensive data set of HPLC pigment, bio-optical, and DNA meta-barcoding observations. Phytoplankton pigment community analyses indicate that 3-5 unique phytoplankton community states are resolvable from HPLC pigment data. Strong statistical relationships between phytoplankton pigment communities and spectral derivatives of phytoplankton absorption coefficients are identified and provide the foundation for the development of a novel bio-optical modeling approach. The bio-optical models reconstruct phytoplankton pigment community dynamics with high skill, and are subsequently applied to an archive of phytoplankton absorption observations to double the size of an approximately monthly HPLC biomarker pigment time series to 20+ years. The extended HPLC pigment time series is analyzed to determine the dominant oceanographic and climate forcings of seasonal to multi-decadal PCC dynamics in the SBC. We find that most phytoplankton groups respond positively to seasonal upwelling, and as a result, are impacted by interannual variability in seasonal upwelling forced by the El Niño Southern Oscillation. On multi-decadal time scales, results from a regional ocean circulation model provide motivation for a data-driven hypothesis that forcing by the North Pacific Gyre Oscillation drives anomalous decadal dinoflagellate blooms by

altering SBC source water advection patterns. Finally, integrating 3.5 years of DNA meta-barcoding observations of SBC protist communities with the HPLC pigment data shows that, despite error in PCC estimates introduced by inter- and intra-group variability in biomarker pigmentation, novel insights into phytoplankton group physiology and function can be gained by integrating PCC methods, and diverse communities of phytoplankton and other protists covary with phytoplankton pigment communities. Overall, this research advances our understanding of large scale forcings and dynamics of SBC phytoplankton communities, and demonstrates a potential path to monitor ecosystem structure and function on unprecedented spatiotemporal scales through ocean color remote sensing.

References

Balint, M., Bahram, M., Eren, A. M., Faust, K., Fuhrman, J. A., Lindahl, B., ... & Tedersoo, L. (2016). Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS microbiology reviews*, *40*(5), 686-700.

Beers J. 1986. Organisms and the food web. *Plankton dynamics of the southern California Bight* **15**: 84–175. Wiley Online Library.

Behrenfeld, M. J., E. Boss, D. A. Siegel, and D. M. Shea (2005), Carbon-based ocean productivity and phytoplankton physiology from space, *Global Biogeochemical Cycles*, *19*(1), doi:10.1029/2004gb002299.

Bidigare, R. R., J. H. Morrow, and D. A. Kiefer (1989), Derivative analysis of spectral absorption by photosynthetic pigments in the western Sargasso Sea, *Journal of Marine Research*, *47*(2), 323-341, doi:10.1357/002224089785076325.

Brzezinski MA, Washburn L. 2011. Phytoplankton primary productivity in the Santa Barbara Channel: Effects of wind-driven upwelling and mesoscale eddies. *Journal of Geophysical Research: Oceans* **116**(C12).

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

Chase A, Boss E, Cetinić I, Slade W. 2017. Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra: toward a global algorithm. *Journal of Geophysical Research: Oceans* **122**(12): 9725–9743. Wiley Online Library.

Countway, P. D., and D. A. Caron. 2006. Abundance and distribution of Ostreococcus sp. in the San Pedro Channel, California, as revealed by quantitative PCR. Appl. Environ. Microbiol. **72**: 2496–2506.

De Vargas, C., S. Audic, N. Henry, and others. 2015. Eukaryotic plankton diversity in the sunlit ocean. Science **348**: 1261605.

Di Lorenzo E, Combes V, Keister JE, Strub PT, Thomas AC, Franks PJ, Ohman MD, Furtado JC, Bracco A, Bograd SJ. 2013. Synthesis of Pacific Ocean climate and ecosystem dynamics. *Oceanography* **26**(4): 68–81.

Di Lorenzo E, Schneider N, Cobb KM, Franks PJS, Chhak K, Miller AJ, McWilliams JC, Bograd SJ, Arango H, Curchitser E. 2008. North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophysical Research Letters* **35**(8).

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*(5374), 237-240, doi:10.1126/science.281.5374.237.

Goericke, R., & Montoya, J. P. (1998). Estimating the contribution of microalgal taxa to chlorophyll a in the field--variations of pigment ratios under nutrient-and light-limited growth. *Marine Ecology Progress Series*, *169*, 97-112. doi:10.3354/meps169097.

Gong, W., Hall, N., Paerl, H., & Marchetti, A. (2020). Phytoplankton composition in a eutrophic estuary: Comparison of multiple taxonomic approaches and influence of environmental factors. *Environmental Microbiology*, *22*(11), 4718-4731. https://doi.org/10.1111/1462-2920.15221.

Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum JR. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**(7600): 465.

Harms, S., and C. D. Winant (1998), Characteristic patterns of the circulation in the Santa Barbara Channel, *Journal of Geophysical Research-Oceans*, *103*(C2), 3041-3065, doi:10.1029/97jc02393.

Henderikx Freitas, F., D. A. Siegel, S. Maritorena, and E. Fields (2017), Satellite assessment of particulate matter and phytoplankton variations in the Santa Barbara Channel and its surrounding waters: Role of surface waves, *Journal of Geophysical Research-Oceans*, *122*, 355–371, doi:10.1002/2016JC012152.

Higgins, H. W., S. W. Wright, and L. Schluter (2011), Quantitative interpretation of chemotaxonomic pigment data, in *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*, vol.**6**, edited by S. Roy et al., pp. 257–313, Cambridge University Press, United Kingdom.

International Ocean Colour Coordinating Group (2014), Phytoplankton Functional Types from Space, *IOCCG Rep. 15*, edited by S. Sathyendranath, Dartmouth, Canada.

Jacox MG, Hazen EL, Zaba KD, Rudnick DL, Edwards CA, Moore AM, Bograd SJ. 2016. Impacts of the 2015–2016 El Niño on the California Current System: Early assessment and comparison to past events. *Geophysical Research Letters* **43**(13): 7072–7080.

Jeffrey, S. W., S. W. Wright, and M. Zapata (2011), Microalgal classes and their signature pigments, in *Phytoplankton Pigments: Characterization,*

Chemotaxonomy and Applications in Oceanography, vol.**6**, edited by S. Roy et al., pp. 3-77, Cambridge University Press, United Kingdom.

Johnson, Z. I., and A. C. Martiny (2015), Techniques for quantifying phytoplankton biodiversity, *Annual Review of Marine Science*, *7*, 299-324, doi:10.1146/annurev-marine-010814-015902.

Mackey M, Mackey D, Higgins H, Wright S. 1996. CHEMTAX-a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Marine Ecology Progress Series* **144**: 265–283.

Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the american Meteorological Society* **78**(6): 1069–1080.

Matson PG, Washburn L, Fields EA, Gotschalk C, Ladd TM, Siegel DA, Welch ZS, Iglesias-Rodriguez MD. 2019. Formation, development, and propagation of a rare coastal coccolithophore bloom. *Journal of Geophysical Research: Oceans* **124**(5): 3298–3316.

Mouw, C. B., Hardman-Mountford, N. J., Alvain, S., Bracher, A., Brewin, R. J., Bricaud, A., ... & Hirawake, T. (2017). A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean. Frontiers in Marine Science, 4, 41.

Ryther, J. H. 1969. Photosynthesis and fish production in the sea. Science 166: 72–76.

Taylor, A. G., M. R. Landry, K. E. Selph, and J. J. Wokuluk. 2015. Temporal and spatial patterns of microbial community biomass and composition in the Southern

California Current Ecosystem. Deep Sea Research Part II: Topical Studies in Oceanography **112**: 117–128.

Van Heukelem, L., and C. S. Thomas (2001), Computer-assisted high-performance liquid chromatography method development with applications to the isolation of and analysis of phytoplankton pigments, *Journal of Chromatography A*, *910*(1), 31-49, doi:10.1016/S0378-4347(00)00603-4.

Venrick, E. L. (2012), Phytoplankton in the California Current system off southern California: Changes in a changing environment, *Progress in Oceanography*, *104*, 46-58, doi:10.1016/j.pocean.2012.05.005.

Vidussi, F., H. Claustre, B. B. Manca, A. Luchetta, and J. C. Marty (2001), Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter, *Journal of Geophysical Research-Oceans*, *106*(C9), 19939-19956, doi:10.1029/1999jc000308.

Wear, E. K., Wilbanks, E. G., Nelson, C. E., & Carlson, C. A. (2018). Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton. *Environmental microbiology*, *20*(8), 2709-2726. https://doi.org/10.1111/1462-2920.14091.

Werdell, P.J., Behrenfeld, M.J., Bontempi, P.S., Boss, E., Cairns, B., Davis, G.T., Franz, B.A., Gliese, U.B., Gorman, E.T., Hasekamp, O. and Knobelspiesse, K.D., 2019. The Plankton, Aerosol, Cloud, ocean Ecosystem mission: status, science, advances. *Bulletin of the American Meteorological Society*, *100*(9), pp.1775-1794. https://doi.org/10.1175/BAMS-D-18-0056.1.

Yeh, Y. C., McNichol, J., Needham, D. M., Fichot, E. B., Berdjeb, L., & Fuhrman, J. A. (2021). Comprehensive single-PCR 16S and 18S rRNA community analysis validated with mock communities, and estimation of sequencing bias against 18S. *Environmental Microbiology*. https://doi.org/10.1111/1462-2920.15553.

II. Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment

Dylan Catlett,[1,*] David A. Siegel,[2]

Journal of Geophysical Research: Oceans **123**(1): 246-264

[1] Earth Research Institute, University of California, Santa Barbara, California
[2] Earth Research Institute and Department of Geography, University of California Santa Barbara, California

*Corresponding author: dsc@ucsb.edu

Abstract

Understanding the roles of phytoplankton community composition in the functioning of marine ecosystems and ocean biogeochemical cycles is important for many ocean science problems of societal relevance. Remote sensing currently offers the only feasible method for continuously assessing phytoplankton community structure on regional to global scales. However, methods are presently hindered by the limited spectral resolution of most satellite sensors and by uncertainties associated with deriving quantitative indices of phytoplankton community structure from phytoplankton pigment concentrations. Here we analyze a data set of concurrent phytoplankton pigment concentrations and phytoplankton absorption coefficient spectra from the Santa Barbara Channel, California, to develop novel optical oceanographic models for retrieving metrics of phytoplankton community composition. Cluster and empirical orthogonal function analyses of phytoplankton pigment concentrations are used to define up to 5 phytoplankton pigment communities as a representation of phytoplankton functional types. Unique statistical relationships are found between phytoplankton pigment communities and absorption features isolated using spectral derivative analysis and are the basis of predictive models. Model performance is substantially better for phytoplankton pigment community indices compared with determinations of the contributions of individual pigments or taxa to chlorophyll *a*. These results highlight the application of data-driven chemotaxonomic approaches for developing and validating bio-optical algorithms and illustrate the potential and limitations for retrieving phytoplankton community composition from hyperspectral satellite ocean color observations.

Introduction

Quantifying variations in phytoplankton biomass and community composition on large spatiotemporal scales is imperative due to their roles in regulating biogeochemical cycles and marine trophic interactions (e.g., Field et al., 1998; Landry, 2002). Several methods have been developed to monitor changes in phytoplankton community structure, including microscopic taxonomy and cell counts, flow cytometric sorting and enumeration, DNA meta-barcoding, and pigment chemotaxonomy via High Performance Liquid Chromatography (HPLC). Each of these methods has strengths and weaknesses (IOCCG, 2014; Johnson and Martiny, 2015), but all require the analysis of field samples limiting the spatial and temporal scales over which phytoplankton community dynamics can be studied. Given the success of ocean color remote sensing for monitoring phytoplankton biomass and net primary production on temporal and spatial scales unattainable with in situ methods, researchers have recently focused on developing methods to monitor changes in phytoplankton community composition (e.g., Moisan et al., 2013; Pan et al., 2011; Torrecilla et al., 2011).

Nearly all remote sensing approaches for assessing phytoplankton community composition have relied on HPLC-derived chemotaxonomy for their validation because of its simple sampling protocol and standardized analysis procedures (e.g., IOCCG, 2014). Chemotaxonomy has inherent uncertainty due to the widespread occurrence of most pigments across many different taxa (see Table 1 below; Jeffrey et al., 2011). Despite this uncertainty, methods have been developed to quantify phytoplankton community composition from measured phytoplankton pigment concentrations, typically in the form of contributions of one or a few groupings based on taxonomy, cell size, or

functional properties ("phytoplankton functional types," or PFT's) to the total concentration of chlorophyll *a*. The two most widely used PFT methods are CHEMTAX (Mackey et al., 1996) and diagnostic pigment (DP) analyses (Uitz et al., 2006). Both make assumptions about the underlying covariations of pigment concentrations, adding uncertainty to the resulting PFT determinations (Higgins et al., 2011). Recently, Empirical Orthogonal Function (EOF) analysis of suites of phytoplankton pigment concentrations has emerged as a promising tool to characterize phytoplankton communities (Anderson et al., 2008; Barrón et al., 2014). Unlike other chemotaxonomic methods, EOF analysis makes no a priori assumptions about the underlying covariance of pigments but this approach has not yet been widely compared to other methodologies.

Many approaches have been developed to retrieve PFTs from multi- or hyperspectral optical measurements (e.g., Bricaud et al., 2007; Li et al., 2013), although multispectral approaches are hindered by their limited ability to resolve phytoplankton-specific absorption features. Forthcoming hyperspectral ocean color remote sensors, such as NASA's Plankton, Aerosol, Cloud, and ocean Ecosystem (PACE) mission (https://pace.gsfc.nasa.gov), will allow for resolution of phytoplankton optical signatures and the potential to develop robust relationships to phytoplankton pigments, taxa, and communities. A promising approach for isolating phytoplankton absorption signatures from hyperspectral ocean color observations is derivative analysis (e.g., Bidigare et al., 1989; Isada et al., 2015). In principle, derivative analysis focuses on the spectral scales at which phytoplankton pigments absorb, thereby isolating phytoplankton-specific absorption signatures in environments where abiotic constituents may be significant (e.g., Craig et al., 2006; Dierssen et al., 2015; Isada et al., 2015; Lubac et al., 2008). However

because many phytoplankton biomarker pigments absorb light at similar wavelengths, it remains difficult to use the isolated absorption features to quantify specific phytoplankton taxa (e.g., Bricaud et al., 2004; Chase et al., 2013).

Here we use an extensive data set of HPLC-derived phytoplankton pigment concentrations and corresponding phytoplankton-specific absorption coefficient derivative spectra from the dynamic coastal environment of the Santa Barbara Channel, California (SBC) to elucidate unique relationships between spectral absorption signatures and phytoplankton pigments. Using a linear modeling approach, we show that concentrations of important biomarker pigments, as well as phytoplankton communities represented by EOF modes, can be accurately modeled. Other widely used PFT metrics such as fractional contributions of individual pigments or taxa to chlorophyll *a* were retrieved with less fidelity. The mechanistic nature of our modeling approach shows that resolution of absorption features across the spectrum is needed to accurately model the concentrations of important biomarker pigments due to their covariance with one another, more ubiquitous pigments, and their corresponding absorption signatures.

Methods

Plumes and Blooms overview

The analyses presented here use data collected as part of the Plumes and Blooms (PnB) program, a time series devoted to understanding ocean color variability in the optically complex Santa Barbara Channel (SBC). PnB has conducted approximately monthly cruises since 1996, sampling at 7 stations along a ~40 km transect crossing the SBC. Methodologies for PnB data products have been described in detail in previous

publications (Anderson et al., 2008; Barrón et al., 2014; Kostadinov et al., 2012; Toole and Siegel, 2001). A brief review of the data used in the present study is presented below.

Discrete seawater samples

Discrete seawater samples were collected from 5-liter Niskin bottles deployed on a rosette for analysis of dissolved inorganic nutrients (NO_3+NO_2, PO_4, and SiO_4, representing nitrate + nitrite, phosphate, and silicate, respectively), biogenic and lithogenic silica (BSi and LSi, respectively), particulate organic carbon (POC), phytoplankton pigment profiles, and determinations of particulate and chromophoric dissolved organic matter absorption coefficients ($a_p(\lambda)$ and $a_g(\lambda)$, respectively). Sea-surface temperature (SST) and salinity were measured using a Seabird 9/11 conductivity-temperature-depth system. The data used in this analysis were measured at or near the sea surface from November, 2005 to December, 2014. Although PnB has been conducting cruises since 1996, quality control problems with HPLC data prior to 2005 preclude its use in this analysis (Hooker et al., 2009).

Nutrient samples were collected in 20-mL plastic scintillation vials and analyzed at UC-Santa Barbara using a Lachat QuikChem 8000 Flow Injection Analyzer (http://www.msi.ucsb.edu/services/analytical-lab/instruments/flow-injection-analyzer). Samples for BSi and LSi analysis were collected on 0.4 μm membrane filters and analyzed using a sodium hydroxide extraction procedure (Krause et al., 2013). POC samples were taken on GF/F filters, stored immediately after collection in liquid nitrogen, and analyzed using a CE440 Elemental Analyzer.

Samples for HPLC analysis of phytoplankton pigments were immediately filtered onto GF/F filters and stored frozen in liquid nitrogen until being analyzed at the Horn Point Laboratory or, later in the time series, at the NASA Goddard Space Flight Center. Both laboratories used the HPLC procedure of Van Heukelem and Thomas (2001). All pigments considered in this analysis are listed in Table 1 along with their respective abbreviations used in the present study, assumed taxonomic value, and known taxonomic distribution. Pigment concentrations below detection limits were assumed to be absent from the sample, and their concentrations were set to zero. The pigments gyroxanthin diester and divinyl chlorophyll b were discarded from all analyses because they were always or almost always (>90% of observations) at or below detection limits, respectively. We also disregarded all chlorophyll a degradation products (chlorophyllide a, pheophorbide a, and pheophytin a).

Samples for $a_p(\lambda)$ were also filtered immediately onto GF/F filters and stored in liquid nitrogen until analysis. $a_p(\lambda)$ was measured using the quantitative filtration technique (Mitchell, 1990) with a Shimadzu 2401-PC spectrophotometer. Following analysis of $a_p(\lambda)$, the filters were extracted in methanol for 48 hours to remove phytoplankton pigments and other organic-soluble material. The detrital absorption coefficient ($a_d(\lambda)$) was then measured using the extracted filter, and was subtracted from $a_p(\lambda)$ to yield the phytoplankton absorption coefficient ($a_{ph}(\lambda)$). The optical path-length amplification factor was determined using natural phytoplankton populations from the SBC (see Barrón et al., 2014 for details). Samples for $a_g(\lambda)$ were collected in glass amber bottles and stored at 4°C for less than 24 hours before filtration through a 0.2 μm filter and analysis on a Shimadzu 2401-PC spectrophotometer. All absorption coefficients were

measured at 1 nm resolution and are considered here across the spectral range 350-700 nm.

Derivative analysis of phytoplankton-specific absorption

Derivative analysis was applied to $a_{ph}(\lambda)$ to identify major absorption signatures and investigate links with phytoplankton pigments and community structure. First and second derivative spectra were computed using a second order finite difference approximation:

(1) $$a'_{ph}(\lambda) = (a_{ph}(\lambda + \Delta\lambda) - a_{ph}(\lambda - \Delta\lambda))/(2 * \Delta\lambda)$$

(2) $$a''_{ph}(\lambda) = (a_{ph}(\lambda + \Delta\lambda) + a_{ph}(\lambda - \Delta\lambda) - 2 * a_{ph}(\lambda))/\Delta\lambda^2$$

where $\Delta\lambda$ is the band separation, or 1 nm in the present study.

A smoothing filter must be applied to the spectral data prior to derivative analysis to reduce the influence of measurement noise. However, the characteristics of the smoothing filter (e.g., type and size) can influence the results of the derivative analysis (Tsai and Philpot, 1998; Vaiphasa, 2006). We investigated four types of smoothing filters (moving average (Tsai and Philpot, 1998), Savitsky-Golay (Tsai and Philpot, 1998), Lowess (Cleveland, 1981), and Hamming window (Harris, 1978)) and selected a smoothing filter by optimizing the linear relationships between selected pigments or pigment sums with their corresponding absorption maxima as identified in the second derivative spectra. This procedure led to the selection of a 15 nm Hamming window filter (see Text S1, Figure S1, Table S1 for procedural details), meaning that $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ are considered here across the spectral range of 358-692 nm. Note however that the optimal smoothing filter may vary between different data sets due to methodological

or instrument differences. Therefore, smoothing filters should be carefully selected when applied to derivative analyses of independent data sets (e.g., Tsai and Philpot, 1998; Vaiphasa, 2006).

Results

Relationships among pigments, community structure, and derivative spectra

Cluster analysis of phytoplankton pigment concentrations

We applied hierarchical cluster analysis using the correlation distance (1-R, where R is the Pearson's correlation coefficient between pigments) and Ward's linkage method to investigate common co-occurrences between phytoplankton pigments (e.g., Latasa and Bidigare, 1998). Preliminary cluster analysis (Figure S2) showed that MVChla, MVChlb, and TChlc co-varied closely with TChla, TChlb, and Chlc1c2, respectively. We therefore did not include MVChla, MVChlb, and TChlc in further analysis since TChla and TChlb have been more widely used in past analyses (e.g., Vidussi et al., 2001), and excluding TChlc appeared to give Chlc1c2 and Chlc3 taxonomic value (discussed below; Figure S2, Figure 1).

Diagnostic pigment (DP) analyses often assume that specific biomarker pigments represent PFTs of different size classes (e.g., Hirata et al., 2011; Uitz et al., 2006). The cluster analysis revealed three phytoplankton pigment clusters separated by linkage distances larger than 1.5 corresponding to the three major phytoplankton size classes (micro-, nano-, and picoplankton) commonly used as a proxy for PFTs (Figure 1). However, there are limitations of using pigment concentrations to derive information on phytoplankton size classes. DP analyses either assume that each DP belongs to a specific

size class or attempt to partition contributions of individual DPs to multiple size classes based on the assumption that the relative abundances of each size class co-varies with TChla (e.g., Hirata et al., 2011; Uitz et al., 2006). For example, different studies have represented TChlb and HexFuco as either pico-, nano-, or both size classes (e.g., Brewin et al., 2010; Hirata et al., 2008; Vidussi et al., 2001). For this data set, TChlb and HexFuco appeared to represent the nanoplankton size class given the large linkage distance between these pigments and the pigments more specific to picoplankton such as Zea and DVChla. Similarly, while Pras can be representative of both the nano- and picoplankton size classes, its position in the cluster analysis suggests it primarily represents picoplankton in this data set.

Within the three broad clusters assumed to represent phytoplankton size classes, five pigment clusters, or communities, were identified representing a diatom-like community, dinoflagellate-like community, mixed nanoplankton-like community, haptophyte-like community, and picoplankton-like community (Figure 1). For simplicity, these pigment clusters are henceforth referred to as diatoms, dinoflagellates, mixed nanoplankton, haptophytes, and mixed picoplankton. The diatom community included Fuco, Chlc1c2, ABCar, and Diadino. Although these pigments are found across many phytoplankton taxa such as bloom-forming dinoflagellates and raphidophytes (Table 1), we assume that this community is dominated by diatoms since they often dominate cell abundances in the SBC (e.g., Anderson et al., 2006; Goodman et al., 2012; Venrick, 1998). TChla is also found within this cluster indicating that diatoms largely control variability in TChla in the SBC, in agreement with past work (Anderson et al., 2008; Venrick, 1998; Venrick, 2012). This also suggests that typical approaches for deriving

PFTs based on the relationships of TChla with DPs will perform poorly. Dinoflagellates, the other major bloom-forming taxa in the SBC (Anderson et al., 2008; Venrick, 1998; Venrick, 2012), are represented by the pigment Perid and also clustered closely to TChla.

The mixed nanoplankton community (Figure 1) consists of TChlb, Neo, Lut, Viola, Allo, and Diato. TChlb, Neo, and Lut are found almost exclusively in green algae (Table 1), and the large linkage distance separating them from Zea, Pras, and DVChla suggests that these pigments primarily represent nano- size green algae in the SBC. The presence of Allo and Viola suggests that red algae in the nano- size class, specifically cryptophytes (via Allo; Table 1), are frequently found together with the green algae. Viola can occur in both red and green algal lineages, but is not known to occur in cryptophytes (Table 1). Given its distance from the green algae-specific pigments, this may suggest the presence of other less prevalent red algae within this community that lack unique biomarker pigments and do not account for a significant portion of the variability in pigments they share with other taxa (e.g., chrysophytes, dictyophytes). The haptophyte community clusters closely with the mixed nanoplankton and consists of HexFuco, ButFuco, and Chlc3. HexFuco occurs almost exclusively in haptophytes while ButFuco and Chlc3 are more widespread in red algae, most notably in the pelagophytes. We assume this pigment group typically represents a haptophyte community due to past observations of blooms of *Phaeocystis spp.* in the SBC (Goodman et al., 2012; discussed below). The final community identified in the cluster analysis is the mixed picoplankton community, which consists of DVChla, Zea, and Pras. These pigments show the weakest overall within-cluster relationships, as well as the weakest relationships with TChla.

The cluster analysis revealed several unexpected divergences in the associations of presumably closely related pigments including Diadino and Diato and Chlc1c2 and Chlc3 (Figure 1). Diato and Chlc3 cluster within the nanoplankton and haptophyte communities while Diadino and Chlc1c2 cluster more closely with the diatoms (Figure 1). The conversion of the light-harvesting Diadino to the photoprotective Diato under high light conditions is well-documented in many phytoplankton taxa (Lohr, 2011, and references therein). The divergence of Diadino and Diato could thus indicate ecological differences between the diatoms and the mixed nanoplankton. In the SBC diatoms proliferate in relatively well-mixed waters with increased macronutrients, chlorophyll concentrations, and light attenuation, while the nanoplankton are more likely to dominate during low-chlorophyll, high-light conditions leading to a greater need for photoprotection (e.g., Anderson et al., 2006; Venrick, 1998; Venrick, 2012). Chlc1c2 and Chlc3 also co-occur in many taxa, including diatoms and haptophytes. Their divergence in the clustering results suggests that the dominant diatom populations in the SBC may lack or have relatively low Chlc3 concentrations (e.g., *Chaetoceros*; Stauber and Jeffrey, 1988), while the dominant haptophyte taxa in the SBC, although unlikely to completely lack Chlc1c2, may have high Chlc3 concentrations relative to Chlc1c2 (e.g., *Phaeocystis*; Zapata et al., 2004). Though we lack direct cell counts in the PnB data set to confirm these hypotheses, both *Chaetoceros* and *Phaeocystis* have been documented as dominant in genera in the SBC and surrounding areas (Goodman et al., 2012; Venrick, 1998; Venrick, 2009).

EOF analysis of phytoplankton pigment concentrations

To further characterize the co-variation among the pigment concentrations, we performed an Empirical Orthogonal Function (EOF) analysis following Anderson et al. (2008). An EOF analysis decomposes the pigment data into a series of orthogonal functions or modes. Each mode is represented conceptually by a set of loadings (presented in Figure 2), which shows the relative contribution of each pigment to that mode, and in space and time by an Amplitude Function (AF; Figure S3) (Anderson et al., 2008; Barrón et al., 2014). Each pigment was mean-centered and divided by its standard deviation prior to the analysis. We included nearly all pigments used in the cluster analysis in the EOF analysis. The exception was Viola, which has limited taxonomic value (Figure 1) and had a disproportionately large influence on the covariance structure of the pigment data. Here, we interpret each pigment EOF mode to represent a distinct phytoplankton community or PFT. Each PFT is thus associated with an AF that describes the relative intensity of that PFT at each point in the data set. Although a detailed analysis of pigment community dynamics is beyond the scope the present work, these PFTs can be correlated to concurrently measured optical (e.g., Figure 4 below) and biogeochemical (e.g., Table S2) parameters to investigate relationships with optical properties and aid in their ecological interpretation.

The first four EOF modes explained 74.5% of the variability in the pigment data and will be discussed here. The five communities identified in the cluster analysis were all represented within these first four EOF modes. Correlations of EOF modes with several environmental parameters are presented in Table S2 and provide additional support for our interpretations of both the EOF modes and pigment clusters. The present

EOF results were in reasonable agreement with the analyses of Anderson et al. (2008) and Barrón et al. (2014) where differences were likely due to the extended pigment suite used here and the different time periods over which the EOF analyses were conducted.

EOF mode 1 was interpreted as an early, mixed bloom community and accounted for 38.7% of the variability in the pigment data set (Figure 2a). The pigments representative of the diatom and mixed nanoplankton communities showed the strongest correlations with this mode, followed by the haptophyte and dinoflagellate communities. Given the significant correlations with cooler temperatures and elevated nutrients (Table S2), this mode is likely to dominate during an upwelling event and shows most taxa responding positively to the introduction of nutrients to the surface ocean. Pigments representative of mixed picoplankton showed weak relationships with this mode, though these relationships varied. Pras was uncorrelated with mode 1, while Zea was weakly but significantly correlated at positive AFs suggesting a role in the mixed bloom community. DVChla was the only pigment that was negatively, although weakly, correlated with mode 1. Although uncorrelated with the other modes by definition, this mode's largest positive AFs occasionally co-occur with large AFs in modes 2, 3, or 4, signifying the emerging dominance of a specific pigment group within the bloom (Figure S3).

Mode 2 is dominated by the mixed nanoplankton and haptophyte communities at positive amplitudes and the diatom community at negative amplitudes, and accounts for 21.4% of the total variance (Figure 2b). The picoplankton also appear to be weakly associated with Mode 2 at positive amplitudes. Thus, mode 2 indicates the alternating dominance of the diatom bloom community and a mixed pico- and nanoplankton community. The diatom bloom community shows a strong correlation with BSi and weak

but significant negative correlations with nutrient concentrations and SST (Table S2), which suggests a well-developed diatom bloom with nutrients mostly drawn down following an upwelling event.

Mode 3 explained 7.7% of the variability in the pigment data and represents the mixed picoplankton at positive AFs and the haptophytes at negative AFs (Figure 2c). This mode showed a weak positive correlation with temperature and weak negative correlations with nutrients, suggesting increased dominance of the picoplankton in warm, nutrient-poor conditions, as expected (e.g., Barrón et al., 2014). The mode 3 AF (Figure S3) occasionally shows pronounced haptophyte blooms in the SBC, with the largest occurring in the spring of 2008. Inner shelf microscopic observations near this time found a bloom of *Phaeocystis spp*. (Goodman et al., 2012), supporting our interpretation of this mode and the corresponding cluster as a haptophyte community.

Mode 4 explained 6.7% of the variability in pigments and primarily represents the dinoflagellates at positive AFs and the diatoms and haptophytes at negative AFs (Figure 2d). The covariance of Fuco with Chlc3 suggests a distinct diatom assemblage from the dominant diatom community identified in mode 2, though we lack sufficient evidence to confirm this and the weak but significant association of the haptophytes with this mode may partially explain the dominance of Chlc3. The significant correlation with BSi implies that this community is primarily composed of diatoms as opposed to other Fuco-containing taxa, and highlights the dominance of diatoms in the SBC as they play significant roles in 3 of the 4 dominant modes of variability in phytoplankton pigments.

Interpretation of the EOF results were roughly supported by results from the CHEMTAX method (Text S2, Figure S4), which aims to partition the contributions of

individual phytoplankton taxa to TChla (Higgins et al., 2011; Mackey et al., 1996). The overall performance of CHEMTAX was poor for this data set. This is likely due to CHEMTAX's underlying assumption that phytoplankton pigment concentrations vary independently from one another. Further discussion regarding the application and results of the present CHEMTAX analysis are provided in Text S2.

Correlations of phytoplankton pigments with derivative spectra

Unique spectral relationships were observed between the concentrations of selected biomarker pigments and values of $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ across specific wavelength ranges as shown in Figure 3. As expected, correlation coefficients were relatively high for spectral bands close to the assumed specific absorption features of individual pigments (Table S1). Biomarker pigments had relationships of varying strength with the absorption signatures of other pigments, and these relationships were generally stronger with the absorption features of the pigments that clustered closely to each respective biomarker pigment. For example, Fuco showed stronger correlations with the absorption signatures of the chlorophyll *c* pigments than all other biomarker pigments. Further, the shapes of the correlation spectra (Figure 3) of pigments found within the same pigment cluster (Figure 1) appeared qualitatively more similar than those of pigments found in different clusters.

We also investigated the relationship between phytoplankton communities diagnosed via EOF analysis and $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ spectra (Figure 4a and 4b, respectively). For each pigment mode, this was done by correlating the value of the AF at each point in the data set (e.g., AF(date, station)) with the value of $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ at

each point in the data set (e.g., $a_{ph}'(\lambda$, date, station)). The pigment EOF modes showed unique patterns of correlations with $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$. Taken together with the co-varying patterns of correlation of specific groups of pigments with $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ (Figure 3) and the cluster analysis (Figure 1), this indicates that absorption features are related to phytoplankton pigment communities rather than individual pigments or taxa. Modes 1, 2, and 4 exhibited significant correlations with $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ across many spectral regions. Mode 1 showed the strongest correlations and is similar to the diatom-associated pigments in its patterns of correlation (Figure 3), although there are clear differences likely due to the association of dinoflagellate and nanoplankton pigments with this mode. Modes 2 and 4 showed weaker correlations with $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ than mode 1. This could be due to similarities in the patterns of correlation observed in the opposing communities that make up these modes. For example, changes in the mode 2 AF signify that the diatom or nanoplankton community is becoming more prevalent at the expense of the other. When the two communities have similar relationships with a particular spectral region, the correlation of the AF with that region will likely be weaker than the correlations observed when each community is considered independently. Mode 3 showed significant correlations in only a few spectral regions, likely due to the smaller roles that picoplankton have on variations in $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ in the SBC (Figure 3). The contributions of the haptophyte and picoplankton communities, as diagnosed via EOF mode 3, to the total phytoplankton assemblage will likely be difficult to infer from $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ from this data set.

Quantitative comparisons of derivative and correlation spectra of pigment communities

Hierarchical cluster analysis was used to investigate the potential to separate pigment communities based on the unique relationships between $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ and pigment concentrations (Figure 5). First, mean values of $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ were computed from subsets of $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ corresponding to observations of the 5% largest concentrations of each pigment considered in the analyses above (hereafter referred to as maximum pigment mean spectra). We then performed cluster analyses on the maximum pigment mean $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ spectra simultaneously using the standardized Euclidean distance metric and the unweighted average distance as a linkage method (Figure 5a). The same $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ spectra were sometimes included in several of the data subsets used to derive each first and second derivative maximum pigment mean spectrum due to the strong covariance amongst the pigments (Figures 1 and 2), confounding this analysis to some extent. We also performed cluster analysis on the pigment-specific correlation spectra presented in Figure 3 using the Euclidean distance metric and the unweighted average distance linkage method (Figure 5b).

The results of both spectral cluster analyses (Figure 5) were in good agreement with the cluster analysis of phytoplankton pigment concentrations (Figure 1). The assumed size-based associations observed in the cluster analysis of pigment concentrations (Figure 1) were similarly separated by the spectral cluster analyses, although the variable associations observed for the Perid maximum mean and correlation spectra resulted in a less clear microplankton cluster. Both spectral cluster analyses indicate that the diatom and dinoflagellate pigments have unique spectral characteristics and should be separable from one another and from the nano- and picoplankton. The

broad nanoplankton and picoplankton pigment clusters remained well-separated from one another and from the diatoms and dinoflagellates in the spectral cluster analyses.

Minor differences were observed between the spectral and pigment cluster analyses (Figures 5 and 1, respectively) within the diatom and mixed picoplankton pigment communities. The spectral associations between Fuco and the more ubiquitous pigments of the diatom pigment community (Chlc1c2, Diadino, ABCar) were weaker than those observed in the pigment concentration cluster analysis (Figure 1). However, the proximity of Fuco to these pigments in the spectral cluster analyses relative to the other biomarker pigments suggests that the spectral absorption signatures of TChla, Chlc1c2, Diadino, and ABCar are strongly associated with diatoms rather than other taxonomic groups in the SBC. Within the mixed picoplankton community, Zea showed stronger spectral associations with Pras despite showing a stronger association with DVChla in the pigment data. Although the mixed picoplankton pigments appear well-separated from the other pigment communities in the spectral cluster analyses, their weak, often insignificant correlative relationships with $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ indicate they will be difficult to model bio-optically in the SBC.

Within the broad pigment community assumed to represent the nano- size class, several differences were observed between the cluster analyses of pigment concentrations (Figure 1) and the corresponding pigment maximum mean and correlation spectra (Figure 5). The correlation and pigment maximum mean spectra of TChlb and Lut were more closely associated with those of HexFuco and ButFuco than the correlation and pigment maximum mean spectra of the other pigments of the mixed nanoplankton community (Allo, Viola, Neo, Diato). The spectral characteristics of all pigments within the broad

nano- size class cluster were separated by linkage distances similar to the within-cluster distances observed for the diatom and mixed picoplankton communities. This suggests that bio-optically discriminating pigments from these two pigment communities will be difficult due to their strong covariance with one another and the resulting similarities in their relationships with spectral absorption signatures.

Modeling phytoplankton pigment communities

Model development with Principal Components Regression

Our goal is to determine the possibilities and limitations of assessing phytoplankton community structure from optical data. Given the unique patterns of correlation between pigments and $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ (Figures 3, 4, and 5b), we modeled phytoplankton pigment concentrations and community metrics as a linear sum of contributions from $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$:

$$p_m = \sum_{i=1}^{N} A_m(\lambda_i) * a'_{ph}(\lambda_i) + B_m(\lambda_i) * a''_{ph}(\lambda_i) + C_m \quad (3)$$

where p_m is the value of the m^{th} modeled pigment concentration or pigment-derived index of phytoplankton community structure, $A_m(\lambda_i)$ and $B_m(\lambda_i)$ are the coefficients of the first and second derivatives of $a_{ph}(\lambda)$ at the i^{th} wavelength λ for pigment concentration or community index m, and C_m is an intercept. The large number of predictor variables and the multicollinearity of the absorption signatures used to predict each p_m necessitated the use of principle components regression to derive each set of $A_m(\lambda)$ and $B_m(\lambda)$ (Massy, 1965). Principal components regression solves the multicollinearity problem by performing principal components analysis (synonymous to EOF analysis) on the predictor variables, $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$. Values of $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ were standardized

to have zero mean and unit variance prior to computing principal components. The AFs of each principal component were used as inputs in a multiple linear regression. Each spectral mode was thus assigned a linear regression coefficient from the multiple linear regression, which was then multiplied by that mode's principal component loadings. These products are summed across all principal components used in the model to generate coefficients for $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$, $A_m(\lambda)$ and $B_m(\lambda)$ (Massy, 1965; see Figure 6 to follow for examples of $A_m(\lambda)$ and $B_m(\lambda)$). More information on the derivation of $A_m(\lambda)$ and $B_m(\lambda)$ is provided in Supplemental Information (Text S3).

Evaluation of significant absorption signatures

Each set of coefficients of the principal components of $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ can be transformed to coefficients of the original derivative spectra (e.g., standardized coefficients shown in Figure 6). The linear modeling approach thus allows for detailed investigations of the most important bands in modeling each p_m. For example, significant bands can be evaluated by examining where the mean values of $A_m(\lambda)$ and $B_m(\lambda)$ were significantly different from zero using their 95% confidence intervals (e.g., Figure 6). Some bands, such as $a_{ph}''(675)$ corresponding to an absorption maximum of chlorophyll *a*, were important in modeling most pigments due to the association of TChla with all phytoplankton taxa (Figure 6). Other bands, such as those near $a_{ph}''(585)$ and $a_{ph}''(640)$ corresponding to absorption maxima of chlorophyll *c*, are more important in modeling those pigments that cluster closely to chlorophyll *c* (e.g., Fuco and HexFuco). Interestingly, the absorption maximum near $a_{ph}''(540)$ was significant in modeling both Fuco and Perid (Figure 6) despite their relatively weak correlative relationships (Figure

3). These results indicate that "communities" of absorption features over the entire spectrum are required to adequately model phytoplankton pigment concentrations and community composition.

Model cross-validation results

The performance of our model was assessed by computing the mean and standard deviation of the mean and median percent errors and the percent bias (e.g., Bracher et al., 2015) of modeled relative to observed p_m, as well as the mean and standard deviation of R^2 and RMSE of the linear relationship between modeled and observed p_m, for each 500-permutation model run (see Text S3). Selected modeling results are presented in Table 2 (see Table S14 for all modeling results). The pigment-derived indices of phytoplankton community structure we attempted to model included pigment to TChla ratios, pigment mode AFs, and contributions to TChla derived from CHEMTAX and several DP analyses based on the methods introduced by Vidussi et al. (2001) and Uitz et al. (2006). A description of how the CHEMTAX and DP community indices were derived is provided in Supplemental Information (Text S4, Table S12).

The linear model (Equation 3) performed well ($R^2 > 0.8$) in retrieving the pigments that dominate the variability in $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$, including TChlb, Fuco, and Perid (Table 2). Reasonable retrievals (R^2 ~0.7) were obtained for HexFuco and Allo, two relatively unambiguous biomarkers of haptophytes and cryptophytes, respectively. These pigments represent four of the five major pigment communities identified in the cluster (Figure 1) and EOF (Figure 2) analyses. Picoplankton pigments were retrieved poorly, with the best retrievals (R^2 ~0.5) achieved for Zea likely due to its stronger

covariance with more readily retrievable pigments as demonstrated in EOF modes 1, 2, and 4 (Figure 2). We were also able to accurately model phytoplankton communities as represented by pigment EOFs ($R^2 > 0.8$), with the exception of EOF mode 3 presumably due to the large influence of the picoplankton on this mode.

The model explained less of the variability in retrievals (as determined by R^2) of fractional contributions of individual taxa or pigments to TChla, including those derived from pigment to TChla ratios, the DP methods of Vidussi et al. (2001) and Uitz et al. (2006), and CHEMTAX (Tables 2 and S14). Retrievals of the CHEMTAX-derived contributions of diatoms and dinoflagellates were reasonably accurate (R^2 ~0.75). However, retrievals of the nanoplankton and haptophytes were poor ($R^2 < 0.6$). The accuracy of CHEMTAX retrievals using this modeling approach relies primarily on the relationships of each taxa with TChla, and demonstrates the limitations of retrieving CHEMTAX-derived community indices in regions where variability in TChla is dominated by one or two taxa. The retrievals of DP-based indices of community structure were also more biased as the mean percent bias was frequently greater than two standard deviations from zero. For example, the model tends to overestimate the fractional contributions of TChlb and Perid to TChla when the observed contributions are low, and underestimate these fractional contributions when they are high. Conversely, the mean percent biases of pigment concentrations and EOFs were generally within 1 or 2 standard deviations of zero, suggesting that bias in these retrievals is less significant.

Our successes in modeling the concentrations of several important biomarker pigments were due to their covariance with other pigments (Figures 1 and 2), and their corresponding absorption signatures (Figures 3, 4, 5, and 6). By taking advantage of this

co-variability, we were also able to predict the relative intensities of several dominant phytoplankton communities as represented by ecologically relevant EOF modes and supported by cluster and CHEMTAX analyses. This again demonstrates the effectiveness of retrieving phytoplankton pigment-derived "communities" rather than individual pigment concentrations or the fractional contributions of individual taxa to TChla.

Discussion

Phytoplankton pigments as a proxy for community structure – benefits of data-driven community analyses

Phytoplankton pigments have been widely used across the world's oceans for assessing community composition and linking community dynamics to ocean optical properties. However, deriving information on community composition through pigment chemotaxonomy requires several assumptions to be made about the taxonomic distribution of the major biomarker pigments (Higgins et al., 2011 and references therein). In particular, the CHEMTAX (Mackey et al., 1996) and DP methods that rely on multiple linear regression (e.g., Uitz et al., 2006; henceforth referred to as "DP analyses") require two additional, often invalid assumptions to derive fractional contributions of taxa to TChla, as discussed below. Further, given the poor correspondence of TChla with primary production, carbon biomass, and cell abundances (e.g., Behrenfeld et al., 2005; Pan et al., 2011; Venrick, 2012) and the strong covariance of the diatom biomarker pigment (Fuco) with TChla in the SBC (Figure 1), the relative contributions of individual taxa to TChla has limited value as an index of community structure.

First, both CHEMTAX and DP analyses assume that pigment to TChla ratios do not vary in time or space nor among species within the same pigment class (Higgins et al., 2011; Mackey et al., 1996; Uitz et al., 2006). However, pigment to TChla ratios have been shown to vary significantly due to intra-specific differences in genetics and physiological responses to environmental variability, as well as between species or genera of the same pigment class (Higgins et al., 2011 and references therein). Although efforts can be made to reduce the uncertainties associated with violating this assumption by partitioning samples based on the environmental conditions in which they were collected when using the CHEMTAX program (e.g., Swan et al., 2016), this process does not account for inter- or intra-specific, genetic, or smaller-scale physiological variations.

The second major assumption applied by both CHEMTAX and DP analyses is that abundances of distinct phytoplankton classes are uncorrelated. This assumption is invalid as concentrations of the major biomarker pigments are strongly and significantly correlated with each other as demonstrated here and by others (e.g., Latasa and Bidigare, 1998; Moisan et al., 2013). Common associations of diverse phytoplankton taxa have also been demonstrated both globally and within the Southern California Bight using microscopic species identifications and counts (Estrada et al., 2016; Reid et al., 1978; Venrick et al., 2009). Additionally, recent global analyses of plankton diversity and community structure based on DNA meta-barcoding revealed unexpected co-variations in many seemingly unrelated microbial taxa and suggested that inter-specific biotic interactions play an important role in controlling community structure (Lima-Mendez et al., 2015). Thus, the pigment independence assumption complicates comparisons with other measures of community composition, degrades the ecological relevance of the

derived phytoplankton communities, and, as demonstrated here, results in less accurate retrievals of phytoplankton community composition from ocean optical properties. The use of data-driven cluster and EOF analyses to define phytoplankton communities relies on these correlations and co-variations among phytoplankton taxa and their respective biomarker pigments, and our modeling approach is successful because it takes advantage of these statistical properties.

EOF analysis requires no assumptions to be made about the underlying pigment data, and has been widely used to characterize the dominant modes of variability of a diverse array of Earth system processes. For example, EOFs are commonly used to describe variability in and provide an index of the phase of several major climate oscillations on regional to global scales, such as the Pacific Decadal Oscillation (e.g., Deser et al., 2010, and references therein). EOF analysis has also been used for decades to characterize microscopic observations of phytoplankton community composition on regional to global scales (Reid et al., 1978; Estrada et al., 2016). In the SBC, Harms and Winant (1998) used EOF analysis to characterize and describe the dominant modes of variability in surface ocean circulation patterns in and around the channel, allowing for the investigation of seasonal and synoptic time-scale succession of these circulation patterns and their drivers. Other examples of EOF analysis applications in the SBC include investigations of the drivers of variability in remote sensing reflectance spectra and spatiotemporal patterns in suspended particle and chlorophyll concentration distributions (Henderikx Freitas et al., 2017; Toole and Siegel, 2001). Thus, data-driven quantitative approaches like EOF analyses provide robust tools to study phytoplankton

community dynamics and their interactions with optical, physical, and geochemical processes.

Le Quéré et al. (2005) defined PFTs so that each PFT would be quantitatively significant on the scales of interest, have a distinct biogeochemical role, be controlled by a unique suite of known environmental parameters, and have an effect on the growth and productivity of other PFTs. These differences are captured to first order by the use of phytoplankton size classes as a proxy for PFTs (e.g., Kostadinov et al., 2010; Le Quéré et al., 2005; Li et al., 2013). The distinction of PFTs by their size is supported by the present work, though our results also highlight the importance of data-driven chemotaxonomic methods for determining typical associations of specific pigments with discrete size classes (Figure 1). Using standard assumptions regarding the associations of biomarker pigments with specific taxonomic groups (Table 1), more refined PFTs were also defined above in the cluster and EOF analyses (Figures 1 and 2), which by definition extract the most quantitatively significant PFTs and allow them to influence the success of other PFTs. Correlative analyses (Table S2) also suggest that the PFTs, or communities, represented by EOF modes are ecologically and biogeochemically relevant. Thus, PFTs defined using data-driven chemotaxonomic approaches may prove to be useful in Earth system models as envisioned by Le Quéré et al. (2005).

We have demonstrated that the EOF analysis offers a viable chemotaxonomic alternative to existing methodologies which are limited due to the assumptions required for their implementation. Data-driven approaches offer a path for retrieving whole phytoplankton pigment communities as opposed to contributions of individual taxa to TChla. It should be noted however that the pigment communities identified in the present

analysis are specific to the SBC. Future work is thus needed to validate the EOF method and modeling approach used here for different regions and on global scales, and to develop and compare novel data-driven chemotaxonomic approaches for characterizing phytoplankton communities.

Derivative analysis and its application to hyperspectral remote sensing observations:

One of our goals is to assess the potential for future satellite missions to retrieve phytoplankton community composition from hyperspectral optical observations. Derivative analysis of hyperspectral observations of $a_{ph}(\lambda)$, $a_p(\lambda)$, and remote sensing reflectance ($R_{rs}(\lambda)$) is an effective tool for isolating phytoplankton absorption features over a wide range of oceanic environments (Bidigare et al., 1989; Craig et al., 2006; Isada et al., 2015; Lubac et al., 2008; Torrecilla et al., 2011). Further, derivative analysis of remotely sensed $R_{rs}(\lambda)$ has the advantage that it removes much of the variability due to the imperfect correction of atmospheric and illumination effects (e.g., Lubac et al., 2008; Philpot, 1991; Tsai and Philpot, 1998). Values of $R_{rs}(\lambda)$ can be modeled as a function of the ratio of the total backscattering coefficient, $b_b(\lambda)$, to the total absorption coefficient, $a(\lambda)$ (e.g., Morel and Prieur, 1977). In order to apply the present modeling approach to hyperspectral $R_{rs}(\lambda)$ spectra, $b_b(\lambda)$ and the non-algal components that contribute to $a(\lambda)$ must contain features that are either known, co-vary with changes in phytoplankton community composition, or are featureless on the spectral scales relevant for resolving phytoplankton absorption signatures. Seawater optical properties are known (e.g., Pope and Fry, 1997; Zhang et al., 2009) and their contribution in a derivative analysis can be

accounted for, leaving the contributions of particulate backscattering and non-algal absorption coefficients to be addressed.

Hyperspectral measurements of the particulate backscattering coefficient ($b_{bp}(\lambda)$) are very rare, particularly from field observation. Phytoplankton culture studies illustrate both inter- and intra-specific variability in the spectral shape of $b_{bp}(\lambda)$ and chlorophyll-specific $b_b(\lambda)$ (Ahn et al., 1992; Bricaud et al., 1983; Bricaud et al., 1988; Stramski et al., 2001; Whitmire et al., 2010). Normalization of $b_b(\lambda)$ to $b_b(400)$ accentuates these spectral differences (Stramski et al., 2001), which suggests that the derivative analysis used here would similarly highlight these differences. It is uncertain how much of these differences would be present when considering the broader taxonomic groupings accessible via phytoplankton pigment observations, making it difficult to speculate as to whether this variability in spectral shape would enhance or hinder the performance of the present modeling approach applied to $R_{rs}(\lambda)$. In the SBC, multi-spectral observations of $b_{bp}(\lambda)$ show that these spectra are largely featureless (Barrón et al., 2014; Kostadinov *et al*., 2012). Further, hyperspectral observations of $b_{bp}(\lambda)$ derived from $R_{rs}(\lambda)$ inversion both in the SBC (Toole and Siegel, 2001) and in a similarly complex coastal environment (Lee and Carder, 2004) suggest that, with some exceptions, $b_{bp}(\lambda)$ is relatively featureless on the spectral scales of phytoplankton absorption. These results suggest that the contributions of $b_{bp}(\lambda)$ to $R_{rs}(\lambda)$ derivative spectra may not significantly affect the applicability of the present approach to $R_{rs}(\lambda)$ in complex coastal environments.

Non-algal materials, such as chromophoric dissolved organic matter (CDOM) and non-algal particles (NAPs), may affect relationships between $R_{rs}(\lambda)$ derivative spectra and phytoplankton pigment communities. The absorption spectra of CDOM and NAPs are

generally similar to one another and featureless as they decrease exponentially with wavelength (e.g., Nelson and Siegel, 2013; Sosik and Mitchell, 1995). However, CDOM absorption spectra from the SBC occasionally have spectral features consistent with absorption due to mycosporine-like amino acids that occur when dinoflagellates dominate the phytoplankton assemblage (Barrón et al., 2014). Further, the effects of CDOM and NAP absorption on the first derivative of $R_{rs}(\lambda)$ could be significant.

As a first order test of the effects of CDOM and NAP absorption on the present modeling approach, we repeated some of the above analysis using the total non-water absorption coefficient ($a_p(\lambda) + a_g(\lambda)$) instead of $a_{ph}(\lambda)$. Identical smoothing filter and derivative computations were used to compute first and second derivatives of the non-water absorption coefficient, and these derivative spectra were used in place of $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ in an identical model cross-validation procedure. Results of these tests are presented in Table S15. Similar fidelity was achieved for retrievals of Perid and Fuco, with mean R^2 values within 3% of those observed with $a_{ph}(\lambda)$. However, the model using non-water absorption explained 8 and 10% less of the variability in TChlb and HexFuco concentrations, respectively, than the model using phytoplankton absorption. Similar results were observed when modeling the amplitudes of the first four pigment EOF modes (Table S15). These results were obtained with no empirical tuning of the modeling approach to non-water absorption coefficient derivative spectra, which may be expected to improve the cross-validation statistics somewhat. Regardless, these results demonstrate the potential to effectively distinguish several major PFTs using remotely sensed hyperspectral $R_{rs}(\lambda)$ if the derivative analysis is able to reliably remove variability arising from atmospheric conditions, illumination intensity, and $b_{bp}(\lambda)$.

Modeling approaches similar to that presented here can also contribute to informing the design of future hyperspectral satellite ocean color sensors by allowing for the reconstruction of wavelength-specific coefficients that can be applied directly to spectral derivatives. Our reconstruction of the coefficients of the spectral derivatives of phytoplankton absorption (Figure 6) suggests that resolving "communities" of spectral absorption features is necessary to bio-optically model phytoplankton community composition. However, depending on the effectiveness of the derivative analysis in removing variability in $R_{rs}(\lambda)$ arising from atmospheric conditions and illumination and $b_{bp}(\lambda)$ variations, the relative importance of individual bands may change when the present modeling approach is applied to hyperspectral $R_{rs}(\lambda)$ observations. This, along with the minor differences in the fidelity of model retrievals when applied to total non-water rather than phytoplankton absorption coefficients, highlights a major pitfall in empirical approaches for the remote sensing of PFTs. For optimal performance, empirical models including those presented here must be tuned for application to new optical properties, oceanic regions, and time periods. Empirical applications of the present and similar modeling approaches to hyperspectral $R_{rs}(\lambda)$ observations are thus necessary to determine optimal spectral band sets for the remote sensing of phytoplankton community dynamics. Despite the limitations associated with empirical bio-optical models, our results indicate that resolving communities of spectral absorption features across the entire spectrum from 350-700 nm is necessary to accurately model phytoplankton pigment communities (Figure 6, Tables 2 and S14).

Thus, derivative approaches applied to remotely-sensed hyperspectral $R_{rs}(\lambda)$ observations hold much potential for assessing phytoplankton community composition on

regional to global scales. Taken together with the fact that derivative analyses may remove much of the variability in $R_{rs}(\lambda)$ arising from poor characterization of atmospheric conditions, this approach may hold great promise for the assessment of phytoplankton community variability for complex coastal environments like the SBC.

Conclusions

We have demonstrated that unique relationships exist between several dominant suites, or communities, of phytoplankton pigments and absorption signatures identified via first and second derivative analysis, and that these relationships can be used to accurately model some metrics of phytoplankton community composition. Our modeling approach is successful because it takes advantage of the covariance of the major biomarker pigments with one another, more ubiquitous pigments, and their respective absorption signatures. This is in contrast to traditional approaches that require that abundances of unique taxa are uncorrelated and that pigment ratios do not vary within a data set. We show that these assumptions are invalid across various spatial and temporal scales and suggest a novel target for in situ validation in the remote sensing of phytoplankton community composition through the concept of phytoplankton pigment communities, which were defined through patterns of correlation with derivative spectra, hierarchical cluster analysis, and EOF analysis. These findings highlight the need for improved quantitative methods for in situ validation of retrievals of community composition from optical properties and the importance of resolving multiple associated phytoplankton absorption signatures to accurately model community composition. Although the present approach was developed using $a_{ph}(\lambda)$, the derivative analysis

provides a promising path for applying this approach to $R_{rs}(\lambda)$ across a wide range of oceanic environments.

Acknowledgements

We acknowledge all past and present members of the Plumes and Blooms team for collecting, processing, and providing support in the use of these data in this work. We also acknowledge Chantal Swan for assistance with CHEMTAX analysis and Crystal Thomas for valuable discussions regarding HPLC analysis. We thank an anonymous reviewer and Collin Roesler for their insightful suggestions that have significantly improved this manuscript. Plumes and Blooms is funded by the National Aeronautics and Space Administration (NASA; NNX11AL94G), and ship time is provided by the National Oceanic and Atmospheric Administration (NOAA) Channel Islands National Marine Sanctuary. This research was also supported by the NASA Biodiversity and Ecological Forecasting program (NASA Grant NNX14AR62A), the Bureau of Ocean and Energy Management Ecosystem Studies program (BOEM award MC15AC00006) and NOAA in support of the Santa Barbara Channel Biodiversity Observation Network. D. Catlett was supported by a NASA Earth and Space Science Fellowship (Grant NNX16AO44HS02). All Plumes and Blooms data products can be accessed from the Plumes and Blooms website (http://www.oceancolor.ucsb.edu/plumes_and_blooms/).

References

Ahn, Y. H., A. Bricaud, and A. Morel (1992), Light backscattering efficiency and related properties of some phytoplankters, *Deep-Sea Research Part a-Oceanographic Research Papers*, *39*(11-12A), 1835-1855, doi:10.1016/0198-0149(92)90002-b.

Aiken, J., Y. Pradhan, R. Barlow, S. Lavender, A. Poulton, P. Holligan, and N. Hardman-Mountford (2009), Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995-2005, *Deep-Sea Research Part II-Topical Studies in Oceanography*, *56*(15), 899-917, doi:10.1016/j.dsr2.2008.09.017.

Anderson, C. R., M. A. Brzezinski, L. Washburn, and R. Kudela (2006), Circulation and environmental conditions during a toxigenic Pseudo-nitzschia australis bloom in the Santa Barbara Channel, California, *Marine Ecology Progress Series*, *327*, 119-133, doi:10.3354/meps327119.

Anderson, C. R., D. A. Siegel, M. A. Brzezinski, and N. Guillocheau (2008), Controls on temporal patterns in phytoplankton community structure in the Santa Barbara Channel, California, *Journal of Geophysical Research-Oceans*, *113*(C4), doi:10.1029/2007jc004321.

Barrón, R. K., D. A. Siegel, and N. Guillocheau (2014), Evaluating the importance of phytoplankton community structure to the optical properties of the Santa Barbara Channel, California, *Limnology and Oceanography*, *59*(3), 927-946, doi:10.4319/lo.2014.59.3.0927.

Behrenfeld, M. J., E. Boss, D. A. Siegel, and D. M. Shea (2005), Carbon-based ocean productivity and phytoplankton physiology from space, *Global Biogeochemical Cycles*, *19*(1), doi:10.1029/2004gb002299.

Bidigare, R. R., J. H. Morrow, and D. A. Kiefer (1989), Derivative analysis of spectral absorption by photosynthetic pigments in the western Sargasso Sea, *Journal of Marine Research*, *47*(2), 323-341, doi:10.1357/002224089785076325.

Bidigare, R. R., M. E. Ondrusek, J. H. Morrow, and D. A. Kiefer (1990), In vivo absorption properties of algal pigments, *Proc. SPIE*, *1302*, 290–302, doi:10.1117/12.21451.

Bracher, A., M. H. Taylor, B. Taylor, T. Dinter, R. Rottgers, and F. Steinmetz (2015), Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations, *Ocean Science*, *11*(1), 139-158, doi:10.5194/os-11-139-2015.

Brewin, R. J. W., S. Sathyendranath, T. Hirata, S. J. Lavender, R. M. Barciela, and N. J. Hardman-Mountford (2010), A three-component model of phytoplankton size class for the Atlantic Ocean, *Ecological Modelling*, *221*(11), 1472-1483, doi:10.1016/j.ecolmodel.2010.02.014.

Bricaud, A., A. L. Bedhomme, and A. Morel (1988), Optical properties of diverse phytoplanktonic species - Experimental results and theoretical interpretation, *Journal of Plankton Research*, *10*(5), 851-873, doi:10.1093/plankt/10.5.851.

Bricaud, A., H. Claustre, J. Ras, and K. Oubelkheir (2004), Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations, *Journal of Geophysical Research-Oceans*, *109*(C11), doi:10.1029/2004jc002419.

Bricaud, A., C. Mejia, D. Biondeau-Patissier, H. Claustre, M. Crepon, and S. Thiria (2007), Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons, *Applied Optics*, *46*(8), 1251-1260, doi:10.1364/ao.46.001251.

Bricaud, A., A. Morel, and L. Prieur (1983), Optical efficiency factors of some phytoplankters, *Limnology and Oceanography*, *28*(5), 816-832.

Chase, A., E. Boss, R. Zaneveld, A. Bricaud, H. Claustre, J. Ras, G. Dall'Olmo, and T. K. Westberry (2013), Decomposition of in situ particulate absorption spectra, *Methods in Oceanography*, *7*, 110-124, doi:10.1016/j.mio.2014.02.002.

Cleveland, W. S. (1981), LOWESS - A program for smoothing scatterplots by robust locally weighted regression, *American Statistician*, *35*(1), 54-54, doi:10.2307/2683591.

Craig, S. E., S. E. Lohrenz, Z. P. Lee, K. L. Mahoney, G. J. Kirkpatrick, O. M. Schofield, and R. G. Steward (2006), Use of hyperspectral remote sensing reflectance for detection and assessment of the harmful alga, *Karenia brevis*, *Applied Optics*, *45*(21), 5414-5425, doi:10.1364/ao.45.005414.

Deser, C., M. A. Alexander, S. P. Xie, and A. S. Phillips (2010), Sea Surface Temperature Variability: Patterns and Mechanisms, *Annual Review of Marine Science*, *2*, 115-143, doi:10.1146/annurev-marine-120408-151453.

Dierssen, H., G. B. McManus, A. Chlus, D. J. Qiu, B. C. Gao, and S. J. Lin (2015), Space station image captures a red tide ciliate bloom at high spectral and spatial resolution, *Proceedings of the National Academy of Sciences of the United States of America*, *112*(48), 14783-14787, doi:10.1073/pnas.1512538112.

Estrada, M., et al. (2016), Phytoplankton across tropical and subtropical regions of the Atlantic, Indian and Pacific oceans, PloS one, 11(3), p.e0151699, doi:10.1371/journal.pone.0151699.

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*(5374), 237-240, doi:10.1126/science.281.5374.237.

Goodman, J., M. A. Brzezinski, E. R. Halewood, and C. A. Carlson (2012), Sources of phytoplankton to the inner continental shelf in the Santa Barbara Channel inferred from cross-shelf gradients in biological, physical and chemical parameters, *Continental Shelf Research*, *48*, 27-39, doi:10.1016/j.csr.2012.08.011.

Harms, S., and C. D. Winant (1998), Characteristic patterns of the circulation in the Santa Barbara Channel, *Journal of Geophysical Research-Oceans*, *103*(C2), 3041-3065, doi:10.1029/97jc02393.

Harris, F. J. (1978), On the use of windows for harmonic analysis with the discrete Fourier transform, *Proceedings of the IEEE*, *66*(1), 51-83. doi:10.1109/PROC.1978.10837.

Henderikx Freitas, F., D. A. Siegel, S. Maritorena, and E. Fields (2017), Satellite assessment of particulate matter and phytoplankton variations in the Santa Barbara Channel and its surrounding waters: Role of surface waves, *Journal of Geophysical Research-Oceans*, *122*, 355–371, doi:10.1002/2016JC012152.

Hirata, T., J. Aiken, N. Hardman-Mountford, T. J. Smyth, and R. G. Barlow (2008), An absorption model to determine phytoplankton size classes from satellite ocean

colour, *Remote Sensing of Environment*, *112*(6), 3153-3159, doi:10.1016/j.rse.2008.03.011.

Hirata, T., et al. (2011), Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types, *Biogeosciences*, *8*(2), 311-327, doi:10.5194/bg-8-311-2011.

Higgins, H. W., S. W. Wright, and L. Schluter (2011), Quantitative interpretation of chemotaxonomic pigment data, in *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*, vol.**6**, edited by S. Roy et al., pp. 257–313, Cambridge University Press, United Kingdom.

Hooker, S., L. Van Heukelem, J. Perl, J. Dolan, and R. Farnbach (2009), CHORS HPLC Uncertainties: Final Report, NASA report, NASA Ocean Biology and Biogeochemistry GSFC Calibration and Validation Office, Available from https://oceancolor.gsfc.nasa.gov/fsg/hplc/chors-hplc/CHORS_Final_Report_Sec.pdf

International Ocean Colour Coordinating Group (2014), Phytoplankton Functional Types from Space, *IOCCG Rep. 15*, edited by S. Sathyendranath, Dartmouth, Canada.

Isada, T., T. Hirawake, T. Kobayashi, Y. Nosaka, M. Natsuike, I. Imai, K. Suzuki, and S. I. Saitoh (2015), Hyperspectral optical discrimination of phytoplankton community structure in Funka Bay and its implications for ocean color remote sensing of diatoms, *Remote Sensing of Environment*, *159*, 134-151, doi:10.1016/j.rse.2014.12.006.

Jeffrey, S. W., S. W. Wright, and M. Zapata (2011), Microalgal classes and their signature pigments, in *Phytoplankton Pigments: Characterization,*

Chemotaxonomy and Applications in Oceanography, vol.**6**, edited by S. Roy et al., pp. 3-77, Cambridge University Press, United Kingdom.

Johnson, Z. I., and A. C. Martiny (2015), Techniques for quantifying phytoplankton biodiversity, *Annual Review of Marine Science*, *7*, 299-324, doi:10.1146/annurev-marine-010814-015902.

Kostadinov, T. S., D. A. Siegel, and S. Maritorena, (2010), Global variability of phytoplankton functional types from space: Assessment via the particle size distribution. *Biogeosciences*, *7*, 3239-3257, doi:10.5194/bg-7-3239-2010

Kostadinov, T. S., D. A. Siegel, S. Maritorena, and N. Guillocheau (2012), Optical assessment of particle size and composition in the Santa Barbara Channel, California, *Applied Optics*, *51*(16), 3171-3189, doi:10.1364/ao.51.003171.

Krause, J. W., M. A. Brzezinski, D. A. Siegel, and R. C. Thunell (2013), Biogenic silica standing stock and export in the Santa Barbara Channel ecosystem, *Journal of Geophysical Research-Oceans*, *118*(2), 736-749, doi:10.1029/2012jc008070.

Landry, M. R. (2002), Integrating classical and microbial food web concepts: evolving views from the open-ocean tropical Pacific, *Hydrobiologia*, *480*(1-3), 29-39, doi:10.1023/a:1021272731737.

Latasa, M., and R. R. Bidigare (1998), A comparison of phytoplankton populations of the Arabian Sea during the Spring Intermonsoon and Southwest Monsoon of 1995 as described by HPLC-analyzed pigments, *Deep-Sea Research Part II-Topical Studies in Oceanography*, *45*(10-11), 2133-2170, doi:10.1016/s0967-0645(98)00066-6.

Lee, Z. P., and K. L. Carder (2004), Absorption spectrum of phytoplankton pigments derived from hyperspectral remote-sensing reflectance, *Remote Sensing of Environment*, *89*(3), 361-368, doi:10.1016/j.rse.2003.10.013.

Le Quere, C., et al. (2005), Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, *Global Change Biology*, *11*(11), 2016-2040, doi:10.1111/j.1365-2468.2005.01004.x.

Li, Z. C., L. Li, K. S. Song, and N. Cassar (2013), Estimation of phytoplankton size fractions based on spectral features of remote sensing ocean color data, *Journal of Geophysical Research-Oceans*, *118*(3), 1445-1458, doi:10.1002/jgrc.20137.

Lima-Mendez, G., et al. (2015), Determinants of community structure in the global plankton interactome, *Science*, *348*(6237), doi:10.1126/science.1262073.

Lohr, M. (2011), Carotenoid metabolism in phytoplankton, in *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*, vol.**6**, edited by S. Roy et al., pp. 113-161, Cambridge University Press, United Kingdom.

Lohrenz, S. E., A. D. Weidemann, and M. Tuel (2003), Phytoplankton spectral absorption as influenced by community size structure and pigment composition, *Journal of Plankton Research*, *25*(1), 35-61, doi:10.1093/plankt/25.1.35.

Lubac, B., H. Loisel, N. Guiselin, R. Astoreca, L. F. Artigas, and X. Meriaux (2008), Hyperspectral and multispectral ocean color inversions to detect *Phaeocystis globosa* blooms in coastal waters, *Journal of Geophysical Research-Oceans*, *113*(C6), doi:10.1029/2007jc004451.

Mackey, M. D., D. J. Mackey, H. W. Higgins, and S. W. Wright (1996), CHEMTAX - A program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton, *Marine Ecology Progress Series*, *144*(1-3), 265-283, doi:10.3354/meps144265.

Massy, W. F. (1965), Principal components regression in exploratory statistical research, *Journal of the American Statistical Association*, *60*(309), 234-256, doi:10.2307/2283149.

Mitchell, B. G. (1990), Algorithms for determining the absorption coefficient of aquatic particulates using the quantitative filter technique (QFT), paper presented at Conf on Ocean Optics 10, Orlando, Fl, Apr 16-18, doi:10.1117/12/21440.

Moisan, T. A. H., J. R. Moisan, M. A. Linkswiler, and R. A. Steinhardt (2013), Algorithm development for predicting biodiversity based on phytoplankton absorption, *Continental Shelf Research*, *55*, 17-28, doi:10.1016/j.csr.2012.12.011.

Morel, A., and L. Prieur (1977), Analysis of variations in ocean color, *Limnology and Oceanography*, *22*(4), 709-722, doi:10.4319/lo.1977.22.4.0709.

Nelson, N.B., and D.A. Siegel (2013), Global distribution and dynamics of chromophoric dissolved organic matter, *Annual Review of Marine Science*, *5*, 447-476, doi:10.1146/annurev-marine-120710-100751.

Organelli, E., A. Bricaud, D. Antoine, and J. Uitz (2013), Multivariate approach for the retrieval of phytoplankton size structure from measured light absorption spectra in the Mediterranean Sea (BOUSSOLE site), *Applied Optics*, *52*(11), 2257-2273, doi:10.1364/ao.52.002257.

Pan, X. J., A. Mannino, H. G. Marshall, K. C. Filippino, and M. R. Mulholland (2011), Remote sensing of phytoplankton community composition along the northeast coast of the United States, *Remote Sensing of Environment*, *115*(12), 3731-3747, doi:10.1016/j.rse.2011.09.011.

Philpot, W. D. (1991), The derivative ratio algorithm - avoiding atmospheric effects in remote sensing, *IEEE Transactions on Geoscience and Remote Sensing*, *29*(3), 350-357, doi:10.1109/36.79425.

Pope, R.M., and E.S. Fry, (1997), Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. *Appl. Opt., 36*, 8710-8723, doi:10.1364/AO.36.008790.

Reid, F. M. H., E. Stewart, R. W. Eppley, and D. Goodman (1978), Spatial distribution of phytoplankton species in chlorophyll maximum layers off Southern California, *Limnology and Oceanography*, *23*(2), 219-226, doi:10.4319/lo.1978.23.2.0219.

Sosik, H. M., and B. G. Mitchell (1995), Light absorption by phytoplankton, photosynthetic pigments and detritus in the California current system, *Deep-Sea Research Part I-Oceanographic Research Papers*, *42*(10), 1717-1748, doi:10.1016/0967-0637(95)00081-g.

Stauber, J. L., and S. W. Jeffrey (1988), Photosynthetic pigments in 51 species of marine diatoms, *Journal of Phycology*, *24*(2), 158-172, doi:10.1111/j.1529-8817.1988.tb04230.x.

Stramski, D., A. Bricaud, and A. Morel (2001), Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community, *Applied Optics*, *40*(18), 2929-2945, doi:10.1364/ao.40.002929.

Swan, C. M., M. Vogt, N. Gruber, and C. Laufkoetter (2016), A global seasonal surface ocean climatology of phytoplankton types based on CHEMTAX analysis of HPLC pigments, *Deep-Sea Research Part I*, *109*, 137-156, doi:10.1016/j.dsr.2015.12.002.

Toole, D. A., and D. A. Siegel (2001), Modes and mechanisms of ocean color variability in the Santa Barbara Channel, *Journal of Geophysical Research-Oceans*, *106*(C11), 26985-27000, doi:10.1029/2000jc000371.

Torrecilla, E., D. Stramski, R. A. Reynolds, E. Millan-Nunez, and J. Piera (2011), Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean, *Remote Sensing of Environment*, *115*(10), 2578-2593, doi:10.1016/j.rse.2011.05.014.

Tsai, F., and W. Philpot (1998), Derivative analysis of hyperspectral data, *Remote Sensing of Environment*, *66*(1), 41-51, doi:10.1016/s0034-4257(98)00032-7.

Uitz, J., H. Claustre, A. Morel, and S. B. Hooker (2006), Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, *Journal of Geophysical Research-Oceans*, *111*(C8), doi:10.1029/2005jc003207.

Vaiphasa, C. (2006), Consideration of smoothing techniques for hyperspectral remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, *60*(2), 91-99, doi:10.1016/j.isprsjprs.2005.11.002.

Van Heukelem, L., and C. S. Thomas (2001), Computer-assisted high-performance liquid chromatography method development with applications to the isolation of and analysis of phytoplankton pigments, *Journal of Chromatography A*, *910*(1), 31-49, doi:10.1016/S0378-4347(00)00603-4.

Venrick, E. L. (1998), The phytoplankton of the Santa Barbara Basin: Patterns of chlorophyll and species structure and their relationships with those of surrounding stations, *California Cooperative Oceanic Fisheries Investigations Reports*, *39*, 124-132.

Venrick, E. L. (2009), Floral patterns in the California Current: The coastal-offshore boundary zone, *Journal of Marine Research*, *67*(1), 89-111, doi:10.1357/002224009788597917

Venrick, E. L. (2012), Phytoplankton in the California Current system off southern California: Changes in a changing environment, *Progress in Oceanography*, *104*, 46-58, doi:10.1016/j.pocean.2012.05.005.

Vidussi, F., H. Claustre, B. B. Manca, A. Luchetta, and J. C. Marty (2001), Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter, *Journal of Geophysical Research-Oceans*, *106*(C9), 19939-19956, doi:10.1029/1999jc000308.

Whitmire, A. L., W. S. Pegau, L. Karp-Boss, E. Boss, and T. J. Cowles (2010), Spectral backscattering properties of marine phytoplankton cultures, *Optics Express*, *18*(14), 15073-15093, doi:10.1364/oe.18.015073.

Zapata, M., S. W. Jeffrey, S. W. Wright, F. Rodriguez, J. L. Garrido, and L. Clementson (2004), Photosynthetic pigments in 37 species (65 strains) of Haptophyta:

implications for oceanography and chemotaxonomy, *Marine Ecology Progress Series*, *270*, 83-102, doi:10.3354/meps270083.

Zhang, X., L. Hu, L., and M.-X. He (2009), Scattering by pure seawater: Effect of salinity. *Optics Express, 17*, 5698-5710, doi:10.1364/OE.17.005698.

Tables and Figure Captions

Table 1. Pigments used in the present analysis and their corresponding abbreviations, assumed taxonomic value, and known taxonomic distribution

Abbreviation	Pigment (pigment sum)	Assumed taxonomic value[a]	Taxonomic distribution[a]
TChla	Total chlorophyll a (TChla = MVChla [including allomers and epimers] + DVChla + Chlorophyllide a)	All taxa	All taxa
TChlb	Total chlorophyll b (TChlb = MVChlb + DVChlb)	All green algae	Cyanobacteria[b], green algae
TChlc	Total chlorophyll c (TChlc = Chlc1c2 + Chlc3)	All red algae	All red algae except rhodophytes and eustigmatophytes
ABCar	Alpha-beta-carotene (β,β-carotene + β,ε-carotene)	All taxa	All taxa
ButFuco	19′-Butanoyloxyfucoxanthin	Pelagophytes	Diatoms, dictyophytes, pelagophytes, raphidophytes, haptophytes, dinoflagellates
HexFuco	19′-Hexanoyloxyfucoxanthin	Haptophytes	Haptophytes, dinoflagellates
Allo	Alloxanthin	Cryptophytes	Cryptophytes, dinoflagellates, chlorophytes
Diadino	Diadinoxanthin	Red algae	Diatoms, bolidophytes, dictyophytes, pelagophytes, xanthophytes, haptophytes, dinoflagellates, euglenophytes
Diato	Diatoxanthin	Red algae	Diatoms, bolidophytes, dictyophytes, pelagophytes, xanthophytes, haptophytes, dinoflagellates, euglenophytes
Fuco	Fucoxanthin	Diatoms	Diatoms, bolidophytes, chrysophytes, dictyophytes, pelagophytes, raphidophytes, haptophytes, dinoflagellates
Perid	Peridinin	Dinoflagellates	Dinoflagellates[c]
Zea	Zeaxanthin	Cyanobacteria, green algae	Cyanobacteria, rhodophytes, diatoms, chrysophytes, dictyophytes, eustigmatophytes, pelagophytes, raphidophytes, all green algae except euglenophytes
MVChla	Monovinyl chlorophyll a	All taxa	All taxa, with the exception of some cyanobacteria
DVChla	Divinyl chlorophyll a	Cyanobacteria	Cyanobacteria
MVChlb	Monovinyl chlorophyll b	Green algae	Cyanobacteria, green algae
Chlc1c2	Chlorophylls c1 and c2 (chlorophyll c1 + chlorophyll c2 + MgDVP[d])	Red algae	Diatoms, chrysophytes, cryptophytes, dictyophytes, pelagophytes, raphidophytes, synurophytes, haptophytes, dinoflagellates
Chlc3	Chlorophyll c3	Red algae	Diatoms, dictyophytes, pelagophytes, haptophytes, dinoflagellates
Lut	Lutein	Green algae	All green algae except euglenophytes
Neo	Neoxanthin	Green algae	Green algae, dinoflagellates
Viola	Violaxanthin	Red and green algae	Diatoms, chrysophytes, dictyophytes, eustigmatophytes, raphidophytes, synurophytes, all green algae except euglenophytes
Pras	Prasinoxanthin	Prasinophytes	Prasinophytes, dinoflagellates

[a]References: Aiken et al. (2009), Jeffrey et al. (2011), Latasa and Bidigare (1998), Moisan et al. (2013), Uitz et al. (2006), and Vidussi et al. (2001).
[b]Here "cyanobacteria" is used to refer to all photosynthetic prokaryotes.
[c]Not all pigmented dinoflagellates contain peridinin.
[d]MgDVP is not likely to have a significant role in the SBC (Crystal Thomas, personal communication, 2016); therefore taxa that contain MgDVP but no chlorophyll c1 or c2 are ignored in this group.

Table 2. Selected model cross-validation results (see Table S14 for complete cross-validation results).

	R^2	Median % Error	% Bias
Pigments	**Mean (Std.)**	**Mean (Std.)**	**Mean (Std.)**
Fuco	0.856 (0.036)	32.6 (5.3)	12.9 (9.5)
Perid	0.887 (0.044)	88.3 (12.7)	169.0 (121.2)
TChlb	0.815 (0.083)	22.5 (2.8)	155.9 (199.2)
HexFuco	0.733 (0.067)	29.9 (3.6)	17.0 (8.2)
Allo	0.721 (0.049)	35.9 (4.4)	144.6 (129.6)
Zea	0.541 (0.098)	37.7 (4.1)	72.6 (63.6)
Community Indices			
EOF Mode 1	0.884 (0.026)	25.3 (3.4)	112.1 (160.5)
EOF Mode 2	0.852 (0.047)	38.0 (4.5)	60.2 (176.6)
EOF Mode 3	0.454 (0.083)	74.8 (7.6)	-91.2 (45.7)
EOF Mode 4	0.809 (0.045)	51.9 (6.7)	-16.8 (52.5)
CHEMTAX - diatoms[a]	0.750 (0.031)	25.5 (2.9)	874.7 (1104.7)
CHEMTAX - dinoflagellates	0.761 (0.059)	56.8 (9.1)	6.8E+14 (1.1E+15)
CHEMTAX - nanoplankton	0.586 (0.054)	25.2 (2.5)	1796.2 (1407.0)
DP, Uitz, local - diatoms[b]	0.676 (0.047)	19.7 (1.9)	18.6 (6.2)
DP, Uitz, local - dinoflagellates	0.695 (0.064)	57.8 (8.9)	141.8 (63.3)

[a]Methods used to derive community composition derived from CHEMTAX and DP analysis are presented in Text S4.
[b]"DP, Uitz, local" indicates fractional contributions of the specified taxa to TChla derived from the locally tuned version of the DP analysis (Uitz et al., 2006).

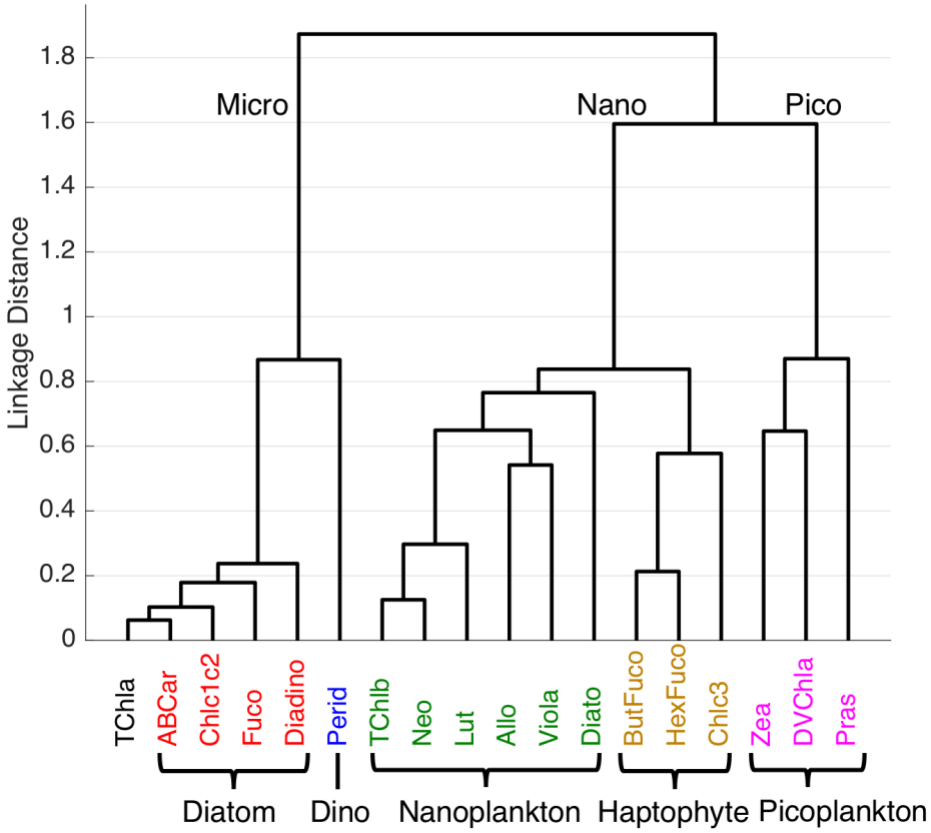

Figure 1. Hierarchical cluster analysis of phytoplankton pigment concentrations using the correlation distance (1-R, where R is Pearson's correlation coefficient) and Ward's linkage method. The assumed size classes are indicated where they diverge in the tree, and the five major pigment communities are highlighted with brackets. Dino is an abbreviation for Dinoflagellates.

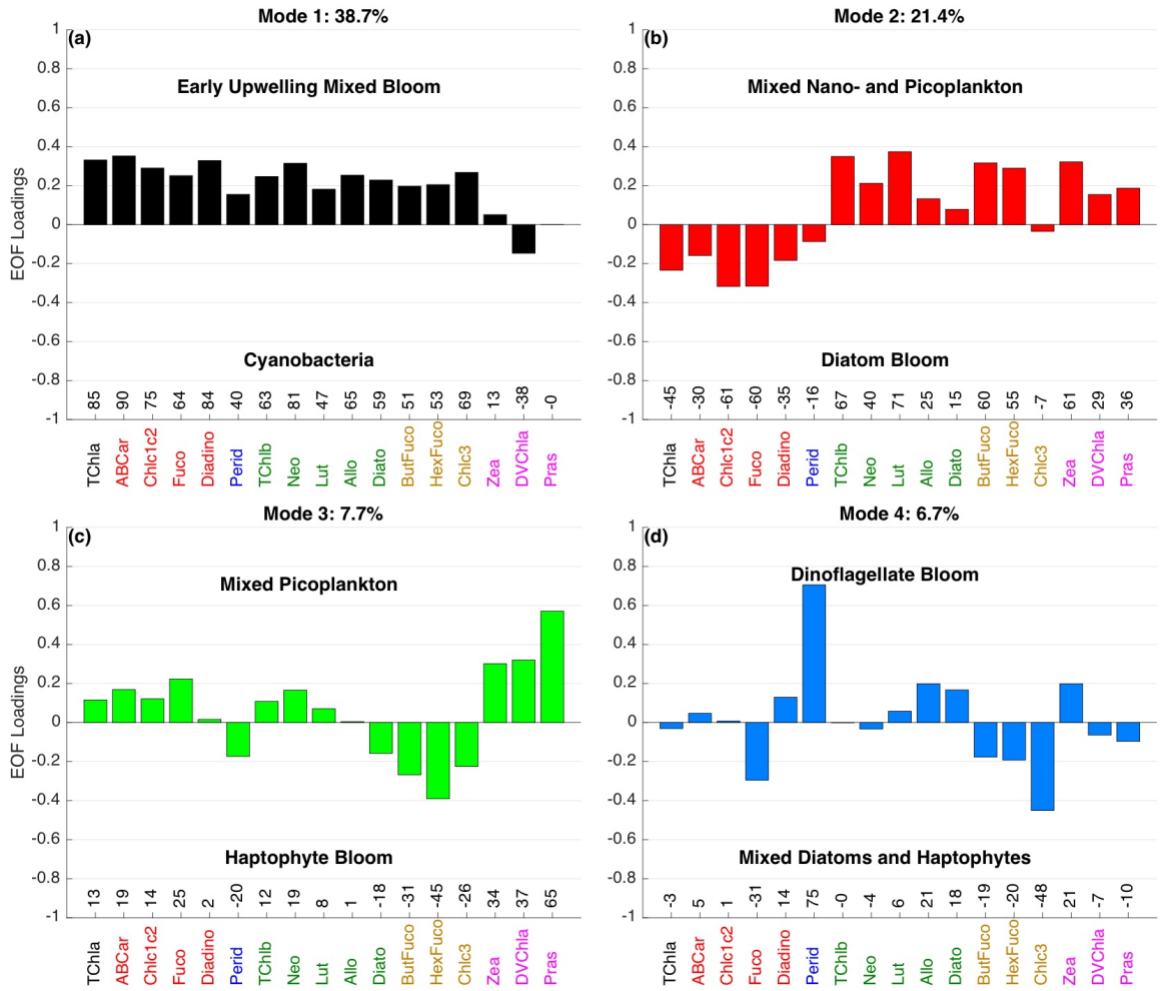

Figure 2. EOF loadings of pigment EOF modes (a) 1, (b) 2, (c) 3, and (d) 4. The mode number and percent variance explained by that mode are listed above each panel, and our interpretations of the taxonomic significance of both the positive and negative loadings of each mode are listed within each panel. Rotated numbers above each pigment represent the correlation coefficient of that pigment with the given mode multiplied by 100. Note that pigments are arranged and color-coded along the x-axis to match Figure 1.

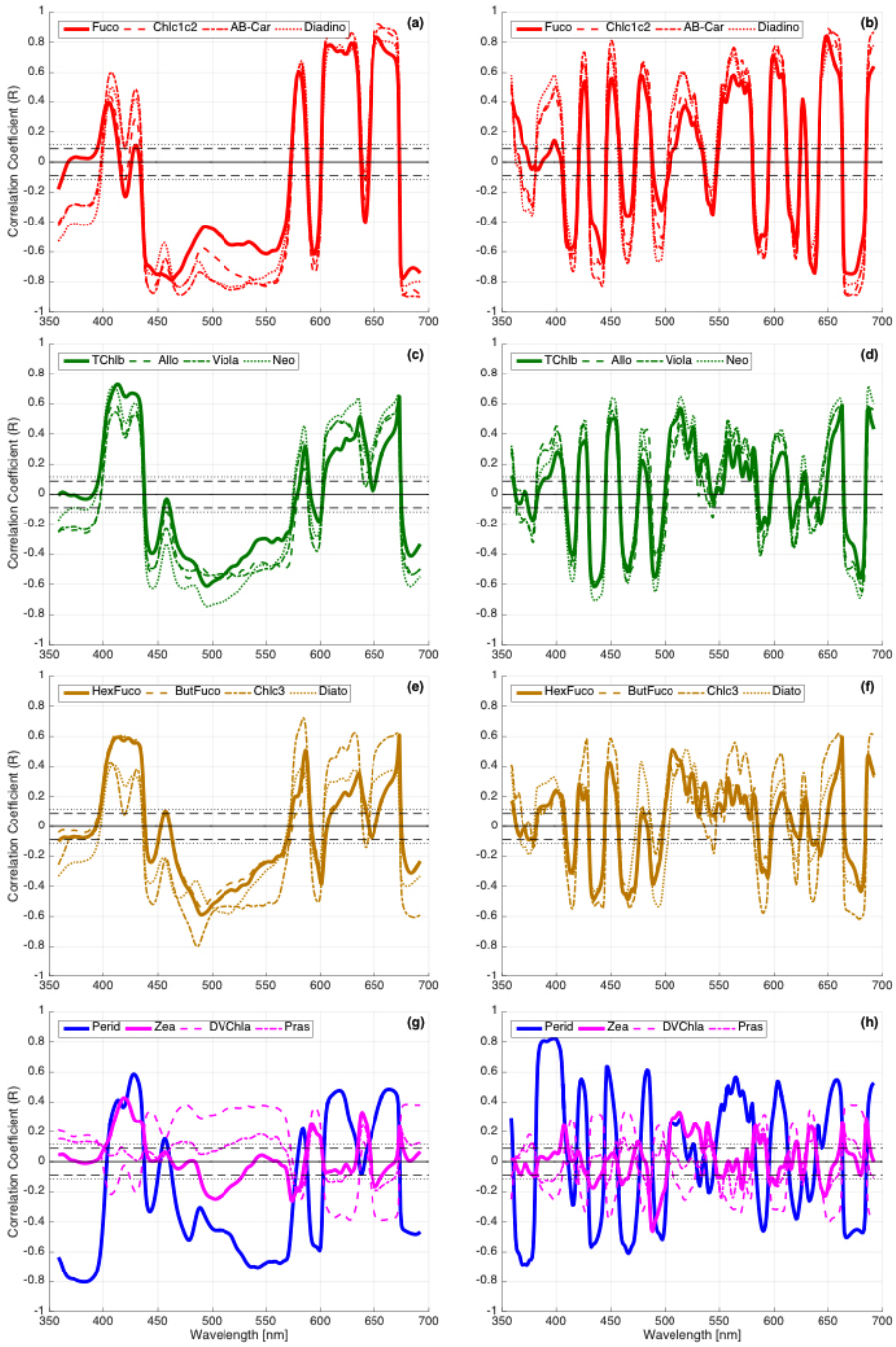

Figure 3. Correlations of selected phytoplankton pigments with $a_{ph}'(\lambda)$ (a, c, e, g) and $a_{ph}''(\lambda)$ (b, d, f, h). Pigments are separated according to the results of the cluster analysis (Figure 1) as (a, b) diatoms, (c, d) nanoplankton, (e, f) haptophytes, and (g, h) dinoflagellates and picoplankton. Note that Lut is not shown for readability, but shows similar patterns to the nanoplankton pigments (c, d), and Diato is included with the haptophytes for readability and because it clusters closely with both the mixed nanoplankton and haptophytes. This data set consists of 491 concurrent measurements of phytoplankton pigment concentrations and $a_{ph}(\lambda)$. The dashed and dotted black lines denote the magnitude of significant correlation coefficients at 95 and 99% confidence, respectively.

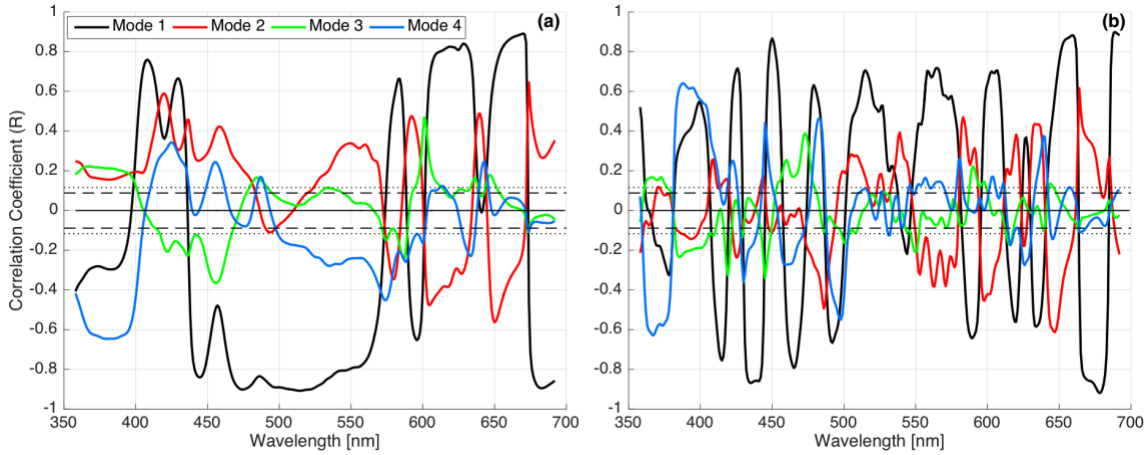

Figure 4. Correlations of the first four pigment EOF modes with (a) $a'_{ph}(\lambda)$ and (b) $a''_{ph}(\lambda)$. The dashed and dotted black lines denote the magnitude of significant correlation coefficients at 95 and 99% confidence, respectively. Line colors correspond to the colors of the bars in Figure 2.

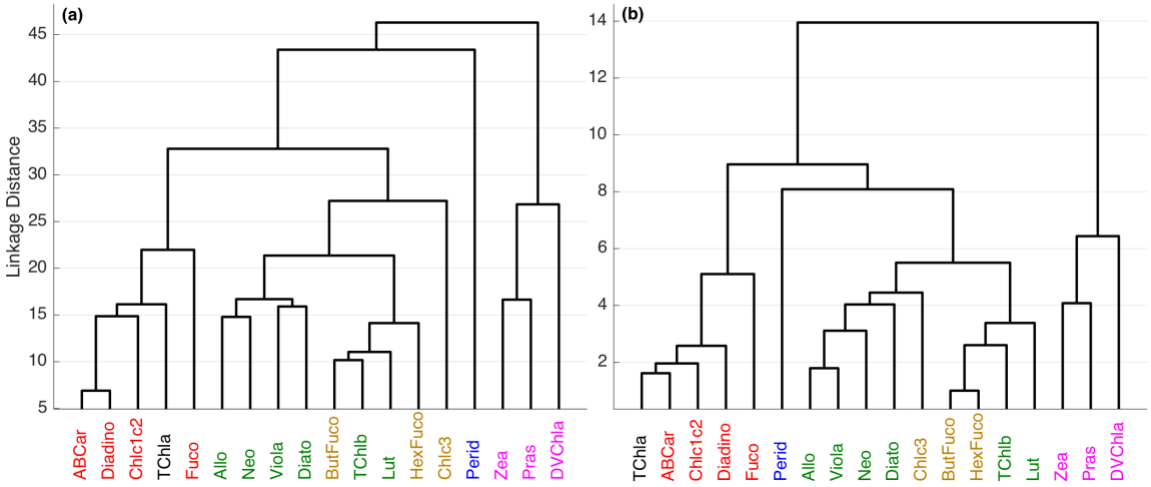

Figure 5. Hierarchical cluster analysis of (a) the mean $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ spectra corresponding to the 5% largest concentrations of each pigment and (b) the correlation spectra (as presented in Figure 3) of each pigment with $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$. (a) uses the standardized Euclidean distance metric while (b) uses the Euclidean distance metric. Both analyses use the unweighted average distance as a linkage method. Pigment labels are color-coded to match those presented in Figure 1.

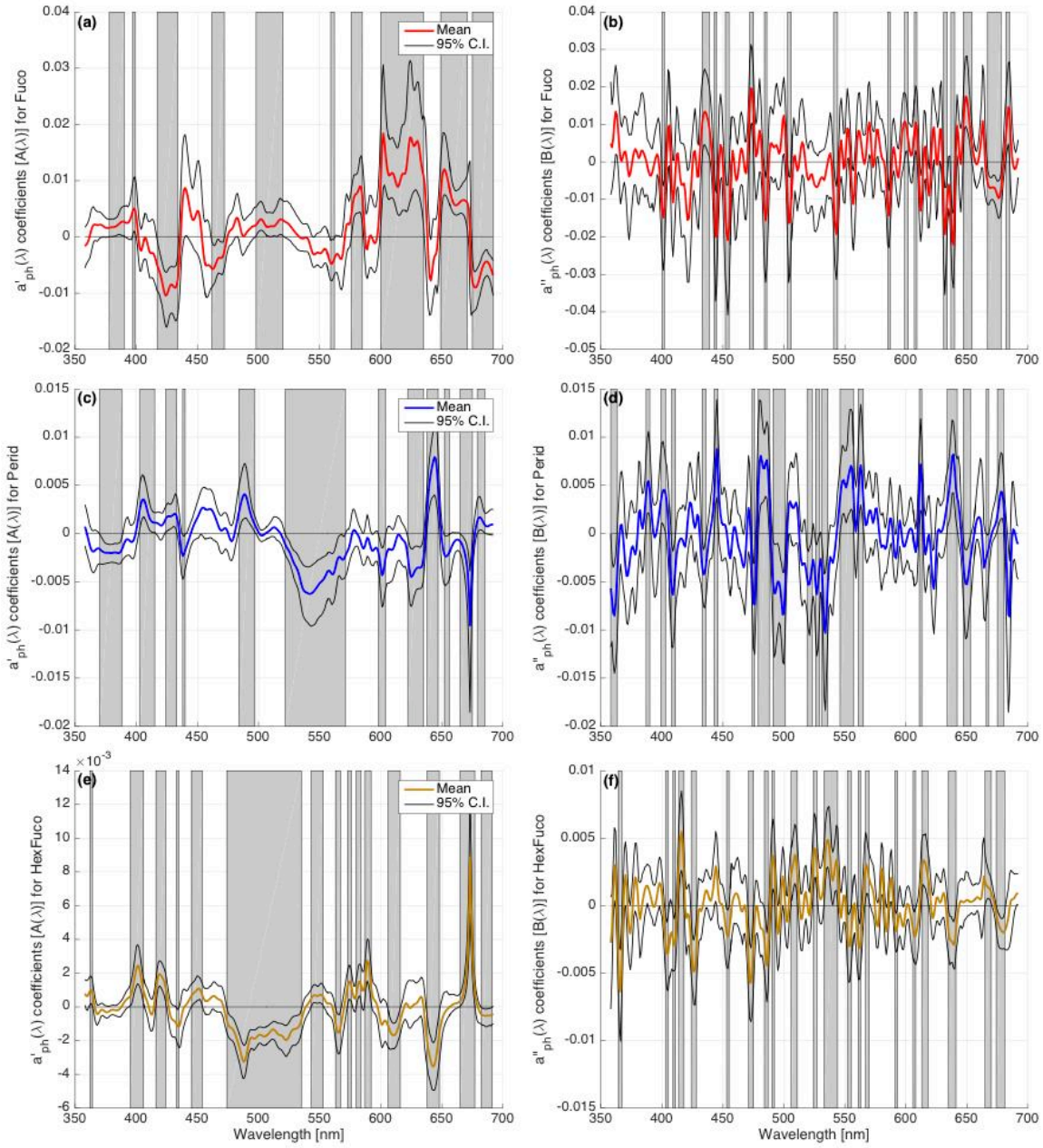

Figure 6. Examples of standardized mean and empirical 95% confidence intervals of the wavelength-specific coefficients of the (a, c, e) first and (b, d, f) second derivative spectra for retrievals of (a, b) Fuco, (c, d) Perid, and (e, f) HexFuco. Shaded areas correspond to spectral regions where the mean coefficients are significantly different from zero. Significant bands with a width of 1 nm are not shaded.

III. Diagnosing seasonal to multi-decadal phytoplankton group dynamics in a highly productive coastal ecosystem

D. Catlett[1]*, D. A. Siegel[1,2], R. D. Simons[1], N. Guillocheau[1], F. Henderikx-Freitas[3], C. S. Thomas[4]

Progress in Oceanography **197**: 102637

[1] Earth Research Institute, UC Santa Barbara, Santa Barbara, CA, USA
[2] Department of Geography, UC Santa Barbara, Santa Barbara, CA, USA
[3] Department of Oceanography, University of Hawai‘i at Manoa, Honolulu, HI, USA
[4] Ocean Ecology Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

*Corresponding author: dsc@ucsb.edu

Abstract

The Santa Barbara Channel, CA (SBC) is a biodiverse marine ecosystem fueled largely by phytoplankton productivity, and the composition of the phytoplankton community influences the magnitude and fates of this productivity. Here, we create a 22-year monthly time series of phytoplankton biomarker pigment concentrations in the SBC by combining 12 years of high performance liquid chromatography phytoplankton pigment concentrations with bio-optical models and 10 additional years of bio-optical observations. The bio-optical models skillfully predict biomarker pigment concentrations representative of five distinct phytoplankton groups (PGs; diatoms, dinoflagellates, chlorophytes, prymnesiophytes, and picophytoplankton) and resolve seasonal responses to the annual upwelling-relaxation cycle for all PGs except the dinoflagellates. Our observations indicate that nanophytoplankton groups respond most rapidly to seasonal upwelling, followed by diatoms, and then by picophytoplankton as the water column stratifies in the summer. A Regional Ocean Modeling System (ROMS) solution is used to relate advection of different source waters to the observed PG dynamics. The ROMS simulation results suggest that, on seasonal time scales, pronounced cross-SBC differences in PG seasonality are related to cross-SBC differences in source waters. El Niño Southern Oscillation events drive interannual variability in the upwelling response of most PGs. On decadal time scales, dinoflagellate blooms are associated with the warm phase of the North Pacific Gyre Oscillation and anomalous advection of Southern California Bight source waters into the SBC. Taken together, our results provide a novel view of phytoplankton community succession in response to seasonal upwelling by considering the dynamics of pico- and nano-phytoplankton and suggest that regional

surface ocean advection plays a substantial role in driving phytoplankton composition in the SBC.

Introduction

The Santa Barbara Channel, CA (SBC, Figure 1) is an exceptionally productive, biodiverse, and well-studied coastal marine ecosystem at the boundary of the relatively cool, productive California Current System (CCS) and the warmer, more oligotrophic Southern California Bight (SCBight) (Beers, 1986; Harms and Winant, 1998; Venrick, 1998; Brzezinski and Washburn, 2011; Henderikx Freitas et al., 2017). Pronounced spatiotemporal gradients in oceanographic features are frequently observed in the SBC due to its location in the "transition zone" between the CCS and SCBight, and phytoplankton blooms are often more intense in the SBC relative to the surrounding region (Harms and Winant, 1998; Venrick, 1998; Brzezinski and Washburn, 2011; Henderikx Freitas et al., 2017). Variations in phytoplankton community composition in the SBC are also dynamic, though under-explored, and are known to impact pelagic food webs and elemental cycling throughout coastal California and the world's oceans (Beers, 1986; Field et al., 1998; Guidi et al., 2016; Lin et al., 2017).

Large spatiotemporal variations in oceanographic properties are often observed in the SBC and result from regional atmospheric and oceanic circulation patterns associated with the annual upwelling-relaxation cycle in combination with the unique geometry of the SBC coastline (Harms and Winant, 1998; Winant et al., 2003; Brzezinski and Washburn, 2011). Point Conception roughly marks the northwest corner of the SBC. Here, the orientation of the California coastline and coastal mountain ranges shifts from

north-south to east-west. This shift creates an upwelling shadow where upwelling winds are strongest on the west side of the SBC near Point Conception and progressively weaken towards the eastern SBC (Harms and Winant, 1998; Winant et al., 2003; Fewings et al., 2015). Over the course of an annual cycle, wind-driven upwelling is strongest during the spring and early summer and to the north and west of the SBC in the southern CCS. Upwelling introduces nutrients to the euphotic zone and allows for the accumulation of phytoplankton, most notably diatoms, in the surface ocean (McPhee-Shaw et al., 2007; Brzezinski and Washburn, 2011; Krause et al., 2013). The persistent wind-driven equatorward flows of the CCS maintain a pressure gradient that drives the poleward flowing California Countercurrent in the nearshore waters off Southern California as well as larger-scale poleward flows at the onset of wind relaxations (Harms and Winant, 1998; Winant et al., 2003; Melton et al., 2009). These circulation patterns result in the entrainment of cold, productive, nutrient-rich waters into the southwestern SBC, and warmer, more oligotrophic waters into the northeastern SBC. While the relative strength of these two circulation patterns can vary on daily to seasonal or longer time scales, the combination of these flows leads to the persistence of a convergent, cyclonic eddy that can further concentrate particles and primary productivity in the central SBC (Harms and Winant, 1998; Brzezinski and Washburn, 2011; Simons et al., 2015). On smaller spatiotemporal scales and in particular on the inner continental shelf, a complex combination of local wind-driven upwelling, internal wave and tide dynamics, freshwater discharge events, and surface gravity waves can significantly influence primary productivity and particle loads in the SBC (Warrick et al., 2004; McPhee-Shaw et al., 2007; Lucas et al., 2011; Henderikx Freitas et al., 2017).

Interannual variations in the physical and biological oceanography of the SBC and surrounding waters are primarily modulated by natural climate oscillations including the El Niño Southern Oscillation (ENSO) (Bograd and Lynn, 2001; Chavez et al., 2002; Venrick, 2012; Jacox et al., 2016), the Pacific Decadal Oscillation (PDO) (Mantua et al., 1997; Chhak and Di Lorenzo, 2007; Di Lorenzo et al., 2013), and the North Pacific Gyre Oscillation (NPGO) (Di Lorenzo et al., 2008; Di Lorenzo et al., 2013). The ENSO exhibits a 3- to 5-year periodicity and has long been recognized as a prominent driver of interannual variations in biological responses to seasonal upwelling in the SBC and surrounding waters. During extreme El Niño events, upwelling winds are suppressed, the water column is anomalously stratified, and warm sea surface temperatures and low phytoplankton biomass are often observed (Bograd and Lynn, 2001; Shipe et al., 2002; Jacox et al., 2014; Jacox et al., 2016). Conversely, La Niña conditions signify an enhancement of seasonal upwelling and a relatively shallow nutricline (Venrick, 2012; Jacox et al., 2016). Several recent studies have demonstrated the role of two dominant modes of North Pacific decadal climate variability, the PDO and NPGO, in driving oceanographic variability in the CCS and SCBight. The PDO is thought to exert a stronger impact on the northern CCS above ~38° N (Di Lorenzo et al., 2008; Di Lorenzo et al., 2013), while the NPGO is associated with low-frequency oscillations in salinity, nutrient concentrations, and phytoplankton biomass in the southern CCS and SCBight (Di Lorenzo et al., 2008; Di Lorenzo et al., 2013). The cold phase of the NPGO (PDO), signified by positive (negative) values of the corresponding statistical index, indicates stronger wind-driven upwelling and enhanced equatorward flows in the southern (northern) CCS, while the warm phase of the NPGO (PDO) is associated with a

relaxation and postponement of seasonal upwelling and enhanced poleward flows in the CCS (Mantua et al., 1997; Di Lorenzo et al., 2008; Di Lorenzo et al., 2013).

Like the physical forcings in the region, phytoplankton communities of the SBC and surrounding waters have been studied for many years and are highly variable across a range of temporal and spatial scales (Allen, 1942; Reid et al., 1978; Goodman et al., 1984; Venrick, 2002; Anderson et al., 2008; Goodman et al., 2012; Venrick, 2012; Taylor et al., 2015; Needham and Fuhrman, 2016). Cell abundances of a given species can vary by orders of magnitude on time scales of days to weeks (Goodman et al., 1984; Bialonski et al., 2016; Barth et al., 2020) and on spatial scales less than one km (Reid et al., 1978; Goodman et al., 2012), and the dominant species within a bloom sampled at a fixed point in space can change daily (Needham and Fuhrman, 2016). Generally, dinoflagellates, prymnesiophytes, and picophytoplankton dominate the phytoplankton community under stratified, low biomass conditions, and in offshore waters of the SCBight and CCS regions (Venrick, 2002; Taylor et al., 2015). Diatoms have repeatedly been shown to dominate cell abundances, carbon biomass, and phytoplankton pigment distributions in the SBC and CCS to the north and west (Venrick, 2002; Anderson et al., 2006; Anderson et al., 2008; Venrick, 2012; Taylor et al., 2015). Along the continental shelf of the SBC and CCS, and more prominently in the SCBight, "red tide" dinoflagellate blooms are frequently observed (Allen, 1942; Gregorio and Pieper, 2000; Barth et al., 2020; Fischer et al., 2020), with longer periods of elevated dinoflagellate abundances observed about once every decade since the early 1900s (Gregorio and Pieper, 2000; Smayda and Trainer, 2010; Fischer et al., 2020). Picoeukaryote blooms dominated by small (< 2 μm) chlorophytes such as *Ostreococcus*, in addition to blooms of the cyanobacterium

Synechococcus, have been documented in the nearshore waters of the SCBight, but have yet to be observed in the SBC (Palenik, 2000; Worden et al., 2004; Countway and Caron, 2006). Finally, blooms of the prymnesiophytes *Phaeocystis sp.* and *Emiliania huxleyi* have been documented in the SBC (Goodman et al., 2012; Wear et al., 2015; Matson et al., 2019), although such observations are less common and, in the case of the recent *E. huxleyi* bloom documented in Matson et al. (2019), unprecedented.

High performance liquid chromatography (HPLC) analysis of phytoplankton pigment concentrations is widely used to assess phytoplankton group (PG) variations (Vidussi et al., 2001; Uitz et al., 2006; Anderson et al., 2008; Kramer and Siegel, 2018). The HPLC method measures the concentrations of ~25 phytoplankton pigments, some of which can be used as "biomarkers" for particular PGs. The benefits of HPLC pigment analysis are the rigorously quality-controlled and standardized analytical procedures (Van Heukelem and Thomas, 2001; Hooker et al., 2010), and the direct links between biomarker pigment concentrations and bio-optical properties that allow for predictions of pigment concentrations from bio-optical observations (Chase et al., 2017; Catlett and Siegel, 2018). As with all methods currently used to quantify phytoplankton community composition, HPLC pigment analysis has limitations and uncertainties; these include ambiguity in the taxonomic identities of many commonly used biomarker pigments, and the complex and variable relationships between biomarker pigment concentrations and cell abundances, carbon biomass, and primary production (Higgins et al., 2011; Jeffrey et al., 2011). Nonetheless, recent work shows that the concentrations of several important biomarker pigments can be modeled from bio-optical observations with high fidelity in the SBC (Catlett and Siegel, 2018), providing an opportunity to create a multi-decadal

biomarker pigment data record for this site. Such large-scale investigations of PG dynamics are of utmost importance as the impacts of climate forcings on PG variations remain poorly understood.

Here, we create an approximately monthly 22-year time series of phytoplankton biomarker pigment concentrations by merging recently developed bio-optical models and 22 years of bio-optical observations with 12 years of HPLC phytoplankton pigment measurements. This time series is used to quantify seasonal to multi-decadal PG variations in the SBC and investigate associations of PGs with oceanographic and climate forcings. Our results demonstrate the dominance of the seasonal upwelling cycle in driving variations of most PGs in the SBC. Using a high-resolution Regional Ocean Modeling System (ROMS) solution, we present evidence that, on seasonal time scales, cross-SBC variability in source waters is linked to spatial variability in PG seasonal cycles. On interannual to decadal time scales, most PGs are impacted by El Niño Southern Oscillation events. Conversely, anomalous decadal dinoflagellate blooms are associated with the warm phase of the North Pacific Gyre Oscillation and anomalous advection of SCBight source waters. This study demonstrates the successful application of a bio-optical model to extend a PG biomarker pigment time series and furthers our understanding of the coupling amongst seasonal upwelling, natural climate oscillations, and advection in driving seasonal to multi-decadal PG variations in the SBC.

Methods

Plumes and Blooms overview

Plumes and Blooms (PnB) has sampled 7 stations approximately monthly on a South-to-North transect in the SBC (Figure 1) since August, 1996 (Otero and Siegel, 2004). Station 7, usually the first station sampled on each PnB cruise, is the Southern-most station and is located on the continental shelf of the Channel Islands in ~75 m of water, while Station 1 is the Northern-most station located on the mainland continental shelf with a water depth of ~45 m (Figure 1). All other PnB stations lie at water depths greater than 200 m. A graphical representation of the coverage of the biomarker pigment data set is shown in Supporting Figure S1. January and February are under-sampled by PnB due to annual ship maintenance, and Station 7 is under-sampled relative to the other PnB stations due to harsher conditions at sea. Multiple PnB cruises were occasionally conducted in the same month, particularly in the first half of the time series. Significant (> 2 months) gaps in the merged biomarker pigment time series (see Section 2.7) occurred from August-October, 2006, March-June, 2010, and July-September, 2018 (Supp. Figure S1).

Due to previously documented analytical issues (Hooker et al., 2010; Barrón et al., 2014), PnB HPLC pigment observations are considered here from November, 2005 to November, 2018 (the same data set considered in Catlett and Siegel, 2018, with the addition of observations from 2015-2018). Bio-optically modeled pigment observations rely on the phytoplankton absorption coefficient ($a_{ph}(\lambda)$), available from April, 1997 to June, 2018. Other PnB data products including potential temperature, salinity, fluorometric chlorophyll *a* concentrations (CHL; see Table 1), and macronutrient concentrations are considered from April, 1997 to November, 2018.

Plumes and Blooms oceanographic observations

To determine associations of PGs with oceanographic processes, we consider potential temperature (henceforth referred to as temperature) and salinity profiles, as well as CHL and dissolved inorganic nutrient concentrations (nitrate, phosphate, and silicate) determined from discrete seawater samples. For most CTD profiles, a Sea-Bird Electronics 911E CTD was deployed on a SBE32C compact carousel. However, from December 2000 to March 2003, a Sea-Bird Electronics SeaCat Profiler CTD was used instead. CTD profiles are considered over the upper 100 m of the water column and are binned to 1 m depth intervals. To quality control the CTD profiles, spurious values of temperature and salinity (temperature > 25 ºC; salinity < 32 psu or > 34.5 psu) were removed and profiles were de-spiked as recommended by the United States Integrated Ocean Observing System (2013). Following de-spiking, 16 temperature and 32 salinity profiles with missing data for > 25% of the depths sampled were discarded.

Discrete seawater samples were collected from 5 L Niskin bottles for analysis of bulk chlorophyll *a* and dissolved inorganic nutrient concentrations. For analysis of chlorophyll *a* concentrations, particles were collected on Whatman GF/F filters via vacuum filtration and immediately frozen and stored in liquid nitrogen. Filters were extracted in 90% acetone overnight and analyzed on a Turner Designs 10AU fluorometer before and after the addition of 2 drops of 1.2 M HCl to determine chlorophyll *a* and phaeopigment concentrations. Samples for dissolved inorganic nutrient concentrations were collected in 20 mL plastic scintillation vials and frozen until analysis with flow injection techniques at the UCSB Marine Science Institute Analytical Lab (Johnson et al., 1985). The detection limits are 0.1 μM for nitrite, 0.2 μM for nitrate plus nitrite, 0.05 μM

for ortho-phosphate, and 0.2 μM for silicate (http://msi.ucsb.edu/services/analytical-lab/seawater-nutrients-fia). Nitrate concentrations are determined by subtracting nitrite concentrations from the total concentration of nitrite and nitrate. Values below detection for phosphate and silicate were set to 0 μM. Where nitrate plus nitrite concentrations were below detection, nitrate values were set to 0 μM. All curated PnB CTD and water sample data analyzed here are publicly available (Catlett et al., 2020a). In all analyses considered here, PnB oceanographic observations are considered from April, 1997 to November, 2018 to match the time period from which HPLC phytoplankton pigment observations are available (see section 2.7 below).

HPLC methods and data-driven PG determinations

Discrete seawater samples for HPLC analysis of phytoplankton pigment concentrations were collected from 5 L Niskin bottles deployed on a rosette and immediately concentrated on GF/F filters by vacuum filtration. Filters were flash-frozen in liquid nitrogen, and stored in liquid nitrogen or at -80°C until HPLC analysis using the method of Van Heukelem and Thomas, 2001. HPLC analysis was conducted at Horn Point Laboratory for all samples collected prior to March, 2011. After March 2011, HPLC analysis was carried out at the NASA Goddard Space Flight Center. Pigment concentrations below the limit of detection were assumed absent from the sample and their concentrations were set to zero. HPLC reporting practices varied over the course of the PnB record. Current practices report pigment concentrations to 0.001 μg L^{-1}. Thus, for data that were reported to four places after the decimal, values were rounded to the nearest thousandth to maintain consistency throughout the HPLC record. The pigments

considered in the present analysis, along with their abbreviations and assumed taxonomic representation, are shown in Table 1.

Although the same HPLC method has been used throughout the PnB record, the limits of quantitation of the method have changed over time due in part to the change in laboratories in 2011 along with changes made to the HPLC slit and step detector settings (which affected analysis of PnB samples collected from February, 2014 to the present). These changes only impact pigments with concentrations that are frequently at, near, or below the limits of detection and quantitation, and can impact assessments of their changes over time. In the PnB data, these pigments include DVChla, Pras, divinyl chlorophyll *b*, and gyroxanthin diester (see Table 1 for a list of pigment names, abbreviations, and assumed taxonomic significance). Divinyl chlorophyll *b* and gyroxanthin diester were not detected in 92.5% and 100% of the PnB samples considered here, respectively, and were not considered in subsequent analysis. Inspection of the dynamics of Pras and DVChla through time (Supp. Figure S2) suggested that measurements of DVChla were reasonably consistent across the two time periods with only minor variations observed that were likely driven by natural variability in the phytoplankton community. However, Pras was detected far more frequently and at higher concentrations after the change in laboratory (Supp. Figures S2 and S3), suggesting analytical artifacts potentially interfered with the measurement of Pras.

The change in HPLC slit and step detector settings for samples collected after February, 2014 improved the signal:noise ratio of the HPLC method by approximately 40%, and therefore is expected to result in more frequent detection of pigments typically present in low concentrations (DVChla and Pras) over the course of the time series. We

directly investigated the effects that the change in HPLC detector settings might have on our interpretations of the dynamics of these two pigments by applying the same, higher limits of quantification measured prior to the change in detector settings, to all DVChla and Pras data collected from 2011 to the present. For each year in the PnB HPLC record (excluding 2005 and 2010 due to insufficient sampling frequency), we then calculated the fraction of PnB stations where Pras or DVChla were not detected (Supp. Figure S3). Very minor changes in the fraction of observations below detection were observed for DVChla with the modified limits of quantification, suggesting the detector change did not substantially impact detection of this pigment in the PnB HPLC data. However, applying the 2011 limits of quantification to all subsequent observations of Pras dramatically increased the fraction of observations where Pras was reported (Supp. Figure S3). Further, from 2009 to 2011 (coinciding with a change in laboratories), the fraction of observations where Pras was not reported decreased nearly two-fold from 76% in 2009 to 43% in 2011. Finally, including Pras in subsequent analyses used for determining PG indices significantly altered our results (Figure 2, Supp. Figure S4). Due to the potential for these results to be driven by analytical artifacts, we thus excluded Pras from further analysis.

To determine the dominant PGs in the PnB HPLC record, we performed hierarchical cluster analysis on the phytoplankton pigment concentration data set using the correlation distance and Ward's linkage method (Latasa and Bidigare, 1998; Catlett and Siegel, 2018; the latter citation is henceforth referred to as CS18). We considered a similar suite of pigments to that used in CS18, but excluded Pras, as well as TChla in order to derive PGs independently of chlorophyll biomass. Five distinct pigment clusters were identified, each of which represent unique PGs that were identical to those

identified in CS18 (Figure 2). Based on the associations of widely used biomarker pigments (Vidussi et al., 2001; Uitz et al., 2006; Jeffrey et al., 2011; Kramer and Siegel, 2018) with each cluster, five representative pigments were selected for further analysis and assumed to represent diatoms (Fuco), dinoflagellates (Perid), chlorophytes (TChlb), prymnesiophytes (Hex), and picophytoplankton (Zea). The concentrations of these biomarker pigments are used as proxies for the pigment biomass of each of these phytoplankton groups (PGs) in the remainder of this paper.

Determinations of spectrophotometric absorption coefficients

Discrete seawater samples were collected on PnB cruises for spectrophotometric determinations of the particulate absorption coefficient ($a_p(\lambda)$). These samples were filtered immediately using GF/Fs and stored in liquid nitrogen until analysis. $a_p(\lambda)$ was measured using the transmittance mode of the quantitative filter technique (Mitchell, 1990; Roesler et al., 2018) on a Perkin-Elmer Lambda 2 spectrophotometer equipped with a Labsphere RSA-PE-20 integrating sphere prior to April, 2003, and on a Shimadzu 2401-PC spectrophotometer equipped with an ISR-2200 integrating sphere from April, 2003 to the present. Following measurement of $a_p(\lambda)$, filters were extracted in methanol for 48 hours and the detrital absorption coefficient, $a_d(\lambda)$, was measured using a procedure identical to that used for $a_p(\lambda)$ on the extracted filter. The beta correction factor was determined empirically using natural phytoplankton communities from the SBC (Guillocheau, 2003). Phytoplankton absorption coefficients ($a_{ph}(\lambda)$) were derived by subtracting $a_d(\lambda)$ from $a_p(\lambda)$ and are considered here from 400-700 nm.

The Perkin-Elmer spectrophotometer occasionally introduced significant noise in estimates of $a_p(\lambda)$ (and by extension, $a_d(\lambda)$ and $a_{ph}(\lambda)$), particularly at shorter wavelengths (< 420 nm). Therefore, additional quality assurance and pre-processing procedures were employed to quality-control estimates of $a_{ph}(\lambda)$. First, the values of $a_d(400)/a_p(400)$ measured on the Perkin-Elmer were compared to the distribution of $a_d(400)/a_p(400)$ measured on the Shimadzu. Anomalously high values of $a_d(400)/a_p(400)$ were occasionally observed in data obtained with the Perkin-Elmer. Therefore, all three component spectra were discarded (104 total observations) when the Perkin-Elmer values were outside three standard deviations of the mean Shimadzu value. An additional 98 IOP determinations from both the Perkin-Elmer (77 observations) and Shimadzu (21 observations) spectrophotometers were discarded for various other reasons (substantial baseline correction errors, highly aberrant spectral shapes, and signatures of incomplete pigment extractions in $a_d(\lambda)$).

Spectral derivative analysis and bio-optical modeling of biomarker pigments

Our goal is to assess patterns and forcings of phytoplankton biomarker pigment dynamics over the course of the 22-year PnB record. However, analytical issues preclude the use of PnB HPLC observations prior to November, 2005 (Barrón et al., 2014), and the integrity of nearly all HPLC samples from 2010 was compromised due to a dewar malfunction. PnB spectrophotometric IOP determinations are available for these time periods, providing a means to extend the time series of the five representative biomarker pigments (Fuco, Perid, TChlb, Hex, and Zea; Figure 2) and TChla to April, 1997. We thus employed a recently developed bio-optical modeling approach to extend the HPLC

time series (CS18). The bio-optical modeling procedure utilizes the first and second derivatives of $a_{ph}(\lambda)$ ($a'_{ph}(\lambda)$ and $a''_{ph}(\lambda)$, respectively) to reliably model biomarker pigment concentrations (CS18). Following CS18, $a_{ph}(\lambda)$ spectral derivatives were calculated using a second order finite difference approximation after the application of a 15 nm Hamming window smoothing filter. Here, we considered values of $a_{ph}(\lambda)$ from 400-700 nm (and thus, spectral derivatives from 408-692 nm after application of the smoothing filter) due to inconsistent sampling of the 350-400 nm spectral range over the course of the PnB record.

Each pigment concentration, p_m, was then modeled as a linear sum of $a'_{ph}(\lambda)$ and $a''_{ph}(\lambda)$:

$$p_m = \sum_{i=1}^{N} A_m(\lambda_i) * a'_{ph}(\lambda_i) + B_m(\lambda_i) * a''_{ph}(\lambda_i) + C_m \quad (1)$$

where $a'_{ph}(\lambda)$ and $a''_{ph}(\lambda)$ are the first and second derivative, respectively, of smoothed $a_{ph}(\lambda)$, $A_m(\lambda)$ and $B_m(\lambda)$ are wavelength-specific coefficients, and $C_m(\lambda)$ is an arbitrary intercept. The empirical derivation of the wavelength-specific coefficients, $A_m(\lambda)$ and $B_m(\lambda)$, for each p_m is described in detail in CS18. Briefly, 500-fold cross-validations of models for each pigment were performed using HPLC observations from November, 2005 to December, 2014 as in CS18. Thus, 500 unique models were developed for each pigment and the average goodness-of-fit statistics for these models are shown in Table 2 and Supporting Table S1. Overall, model performance was similar to that seen in CS18 and the 5 dominant biomarker pigments and TChla were consistently modeled with high fidelity in the cross-validation exercise. The concentrations of each of the five biomarker pigments were thus modeled at all PnB stations where $a_{ph}(\lambda)$ was available. Each p_m was modeled according to equation 1 using the ensemble mean of the 500 models (consisting

of 500 $A_m(\lambda)$, $B_m(\lambda)$, $C_m(\lambda)$; see Supp. Figure S5 for mean +/- 95% confidence intervals of coefficients used for each p_m) determined during the 500-fold cross-validation. In order to maintain consistency with the HPLC data set, modeled pigment concentrations less than 0.0005 μg L^{-1} were replaced by 0 μg L^{-1}, and modeled concentrations greater than or equal to 0.0005 μg L^{-1} but less than 0.001 μg L^{-1} were rounded to 0.001 μg L^{-1}.

Additional independent model validations

The bio-optical models employed here may be particularly susceptible to two sources of uncertainty. First, the use of different spectrophotometers throughout the PnB record may bias the modeled pigment dynamics since different instruments introduce different degrees of noise into spectral IOP measurements, which can then be accentuated by spectral derivative analysis. Second, the bio-optical models used here are formulated empirically, and so may be susceptible to overfitting to the training data. This may result in high uncertainties when the models are extrapolated to new time periods and environments.

To ensure pigment concentrations were modeled consistently across the different spectrophotometers, we compared fluorometric chlorophyll *a* concentrations (CHL) to TChla measured directly by HPLC ($TChla_{HPLC}$), modeled TChla derived from the Shimadzu 2401-PC $a_{ph}(\lambda)$ ($TChla_{Shim}$), and modeled TChla derived from the Perkin-Elmer Lambda 2 $a_{ph}(\lambda)$ ($TChla_{PE}$; Figure 3). CHL concentrations have been assessed using the same method throughout the PnB record. Monovinyl and divinyl chlorophyll *b* and chlorophyll *c* bias CHL determinations (Trees et al., 1985), so these comparisons are not expected to result in perfect agreement since the ratios of chlorophylls *b* and *c* to

total chlorophyll *a* are not stable across samples. Nonetheless, comparing the three different indices of total chlorophyll *a* concentrations with CHL concentrations is expected to reveal any major biases in modeled pigment concentrations that may have been introduced by either spectrophotometer. As expected, all three chlorophyll *a* concentration metrics showed strong, statistically significant, linear relationships with CHL, with the strongest relationship observed for $TChla_{HPLC}$ (Figure 3A). The TChla residuals for all regression analyses were normally distributed and generally low in magnitude, with no obvious bias in any of the regression analyses (Figure 3B-D). While the slopes of all three regressions were significantly different from one another, most values of $TChla_{Shim}$ and $TChla_{PE}$ were within the 95% prediction intervals computed for the $TChla_{HPLC}$-CHL regression (Figure 3A). Taken together, these results suggest that the bio-optical models used here provide consistent estimates of chlorophyll *a* concentrations across the two spectrophotometers. We assume these results can be extrapolated to the modeled concentrations of the biomarker pigments considered here.

Additional independent model validations were performed to verify that the bio-optical models were not overfit to the training data used in the cross-validation exercise above. Since the cross-validations were performed with HPLC observations from 2005-2014, for this test we validated modeled concentrations of TChla and each of the five biomarker pigments against concurrent PnB HPLC observations from 2015 to 2018 (Figure 4). Model performance over this time period was excellent for the diatom and chlorophyte biomarkers ($R^2 > 0.8$, slope ~1; Figure 4B and 4D), with reasonable retrievals found for the prymnesiophyte biomarker ($R^2 > 0.7$, slope = 1.07; Figure 4E). While the model was able to reliably capture dinoflagellate "blooms" (Perid > ~0.5 μg L^-

[1] was generally modeled as such; $R^2 = 0.91$), it consistently underestimated Perid concentrations when HPLC measured concentrations were high (slope > 1.5), and performed poorly when they were low (< ~0.3 μg L^{-1}; Figure 4C). Similar to the results of the cross-validation exercise, the model was able to explain ~50% of the variance in picoplankton pigment concentrations with minimal bias (slope = 1.04; Figure 4F). The reduced fidelity in modeled Zea concentrations is likely due to the low concentrations typically observed in the SBC relative to the other biomarker pigments, and its resulting small contributions to spectral absorption in this data set. Repeating this analysis with all available HPLC pigment observations from 2005-2018 (including the training data; Supp. Figure S6) and further inspection of the space-time distribution of residual errors over this time period (Supp. Figure S7) confirmed that the bio-optical models used here skillfully predict biomarker pigment concentrations across a wide range of oceanographic conditions and phytoplankton community states.

The "merged" pigment data set

All bio-optically modeled pigment concentrations lacking corresponding HPLC observations were merged with all HPLC observations to create a ~22-year, approximately monthly time series of TChla and the five major biomarker pigments at each of the seven PnB stations. Pigment concentrations from November, 2005 to November, 2018 were predominantly measured by HPLC. Pigment concentrations from April, 1997 to October, 2005, and from February to November of 2010 were estimated by the bio-optical models. The application of the bio-optical models extended the HPLC data set of 758 observations over 12 years to 1393 observations spanning roughly 22

years. This data set provides unprecedented spatiotemporal coverage of PG dynamics in the SBC. All phytoplankton pigment data presented here are publicly available (Catlett et al., 2020a).

Ancillary data

We consider additional data beyond that available from PnB in the present analysis. These data include indices of the dominant modes of North Pacific climate variability, simulated surface ocean circulation patterns within and around the SBC, and microscopic counts of several species of diatoms and dinoflagellates at Stearns Wharf on the mainland shelf of the SBC.

Climate oscillation indices

Indices of climate oscillations used here include NOAA's Multivariate ENSO Index (MEI; https://psl.noaa.gov/enso/mei/) (Wolter and Timlin, 1993), the Southern Oscillation Index (SOI; https://www.ncdc.noaa.gov/teleconnections/enso/indicators/soi/data.csv), the Pacific Decadal Oscillation index (PDO; https://www.ncdc.noaa.gov/teleconnections/pdo/data.csv) (Mantua et al., 1997), and the North Pacific Gyre Oscillation index (NPGO; http://www.o3d.org/npgo/npgo.php) (Di Lorenzo et al., 2008). El Niño conditions are indicated by positive (negative) values of the MEI (SOI), while the warm phase of the PDO (NPGO) is indicated by positive (negative) values. All climate indices are presented as provided by their maintainers and

all include monthly values except for the MEI, where bimonthly means considering the preceding month's values are used (Wolter and Timlin, 1993).

Identifying PnB source waters with ocean circulation and particle tracking models

To determine the source waters of the PnB transect, a three-dimensional ocean circulation and Lagrangian particle tracking model was used. The ocean circulation model is a high-resolution Regional Ocean Modeling System (ROMS) solution for the SCBight region (Dong et al., 2009; Dong et al., 2017). The model domain is 674 km by 514 km with 1 km horizontal resolution and 42 vertical levels, and covers the California coast from Point Sur to the southern border with Mexico (Supp. Figure S8). The 1-km grid was nested from a larger 4-km grid that covers the U.S. West Coast (Dong et al., 2017). Our analyses are based on a 10-year ROMS hindcast solution for the years 2004 to 2013, which is stored as hourly offline solutions. As detailed in Dong et al (2017), the ROMS surface boundary conditions came from hourly Weather Research and Forecasting (WRF) products, and lateral boundary conditions from daily HYbrid Coordinate Ocean Model (HYCOM) global oceanic reanalysis products. The ROMS has been shown to accurately reproduce long-term means and seasonal and interannual variability of SCBight circulation (Dong et al., 2009, 2017) and resolve mesoscale features, such as eddies and upwelling, in the SBC (Dong et al., 2011; Simons et al., 2015).

The source waters of the PnB transect were identified using a Lagrangian particle tracking model driven by the ROMS-simulated flow fields (Carr et al., 2008; Simons et al., 2013) and has been used extensively in the SBC and SCBight (Mitarai et al., 2009; Simons et al., 2015). Modeled particle trajectories have also shown good correspondence

with surface drifter observations (Ohlmann and Mitarai, 2010). Using the offline ROMS flow fields, surface-following particles were projected backwards in time using the hourly ROMS output. In order to accurately capture the mesoscale circulation simulated by the ROMS, the PnB transect is represented by 34 particle release locations, approximately 1-km apart, that span PnB sample locations (Figure 1). Over the 10-year hindcast, particles were released daily from each of the 34 points along the PnB transect and tracked backwards in time along the water's surface for 15 days, for a total of 124,000 particles released and tracked over this time period.

We estimate the relative influence of SBC (local), SCBight, and southern CCS source waters on the PnB transect by computing the number of particles that originated from each of four "origin boxes" (West, Center, East, and South; see Figures 1 and 13). The origin box boundaries were selected based on qualitative evaluation of previous observations of surface ocean circulation and satellite observations of long-term mean sea-surface temperature (SST) and chlorophyll *a* concentrations in and around the SBC (Harms and Winant, 1998; Henderikx Freitas et al., 2017). The location of the lines demarcating the West, Center and East origin boxes were chosen such that the dominant circulation patterns would result in the transport of SCBight waters into the eastern entrance of the SBC and of southern CCS waters into the western entrance of the SBC on the advection time scales considered here. Particles originated from the South origin box only rarely, and are largely ignored here. The boundaries of the West and East origin boxes were also placed such that these two origin boxes would largely avoid the steepest east-west gradients in long-term average SST and chlorophyll *a* concentrations shown in Henderikx Freitas et al. (2017).

Daily time series of the number of particles originating from each origin box for each release point were constructed for 5-, 10- and 15-day advection times. Each daily time series was then binned by month. Monthly time series for individual PnB stations (Figure 1) were determined by binning the monthly time series of the four (for PnB station 1) or five (for PnB stations 2-7) closest release points to each PnB station, and another time series was created for the PnB transect by binning the monthly time series of all 34 release points. We consider the proportion of particles originating from the West, East, and Center origin boxes as proxies for the relative magnitude of advection of CCS, SCBight, and SBC source waters in each month of these time series, respectively. The proportion of particles originating from the South origin box was small (see Section 4.3 below), so these source waters are largely ignored in our discussion of these results. In section 4.3 we focus our discussion on results from 10-day hindcasts; qualitatively similar patterns (with expected differences in magnitudes of the proportion of particles from each origin box) were observed for 5- and 15-day advection times and those results are shown in Supporting Figures S9 and S10.

Additional phytoplankton group observations

To aid our discussion of seasonal succession in SBC PGs (see Section 4.4. and Supp. Figure S12 below), we consider observations of the abundances of several prominent diatom and dinoflagellate species at Stearns Wharf on the mainland shelf of the SBC. The Southern California Coastal Ocean Observing System (SCCOOS) Harmful Algal Bloom Monitoring Program provides observations of the abundances of several phytoplankton species (*Akashiwo sanguinea*, *Alexandrium sp.*, *Dinophysis sp.*,

Lingulodinium polyedra, *Prorocentrum sp.*, *Ceratium sp.*, and *Cochlodinium sp*, *Pseudo-nitzschia sp.*) via the National Oceanic and Atmospheric Administration's ERDDAP data portal (https://erddap.sccoos.org/erddap/tabledap/HABs-StearnsWharf.html). Phytoplankton species abundances are determined via microscopic identification and enumeration and are available approximately weekly. These data were retrieved on July 6, 2020 and are considered here from June 30, 2008 to February 10, 2020. Seasonal cycles for total dinoflagellates and for *Pseudo-nitzschia sp.* were computed from monthly mean time series of log-transformed weekly species counts. One cell mL^{-1} was added to both *Pseudo-nitzschia sp.* and total dinoflagellate counts to prevent undefined log-transformed values.

Results

Overview of SBC PG dynamics

Each of the five PG biomarker pigment concentrations displayed unique spatiotemporal dynamics over the course of the 22 year time series (Figure 5). The diatom biomarker pigment, Fuco, was found at higher overall concentrations than the other four biomarker pigments (Table 3), and largely mirrored the TChla patterns (Table 3, Figure 5A and 5B). While Perid was most often found at much lower concentrations than the diatom biomarker pigment, and comparable concentrations to the prymnesiophyte and chlorophyte biomarker pigments, the dynamic range and variance in Perid was high (Table 3). Dinoflagellate biomarker pigment concentrations (Perid) showed the weakest correlations with other biomarker pigment concentrations (Table 3). The prymnesiophyte and chlorophyte biomarker pigment concentrations, Hex and TChlb,

were strongly correlated with one another, though Hex was found at significantly higher concentrations on average (Table 3; two-sample t-test, $p < 0.001$). The picophytoplankton biomarker pigment Zea was correlated with TChlb and Hex (Table 3), and was found at significantly lower concentrations (two-sample t-test, $p < 0.001$ in all comparisons) and displayed the smallest dynamic range of the five biomarker pigments (Table 3).

Fuco showed a clear seasonal cycle with blooms in the spring and annual minima in the fall (Figure 5B and 6B). The magnitude of spring diatom blooms increased later in the time series, beginning in 2008. Spatial variations in diatom pigment concentrations were also apparent, with higher concentrations often observed at the southern PnB stations relative to those closer to the mainland coast. Large diatom bloom events were observed in the PnB record in 2002, 2008, 2010, 2016, and 2017, while relatively low Fuco concentrations were observed consistently from 1997-2001. Conversely, dinoflagellate concentrations periodically increased at PnB station 1, but were typically observed at low concentrations at all other stations (Figure 5C). Anomalous SBC-wide dinoflagellate blooms were observed in late 2003 and early 2006, and again in late 2017 and early 2018. Years with pronounced dinoflagellate blooms typically coincided with an apparent suppression of diatom blooms.

Of the five PGs investigated, the biomarker pigments for prymnesiophytes and chlorophytes, Hex and TChlb, showed the most similar dynamics to one another (Figure 5D and 5E, Table 3). These PGs were typically at their highest concentrations in the winter and early spring, but also sporadically increased at other times of year. Picophytoplankton biomarker pigment concentrations (Zea) generally followed the

opposite pattern of that observed in Fuco, with relatively high concentrations observed in the summer and fall and at the northern PnB stations (Figure 5F).

PG seasonal cycles

Mean seasonal cycles for each PG were calculated by computing monthly mean pigment concentrations after averaging by cruise (spatially) and then, where multiple cruises occurred in the same month, by month (Figure 6). The strongest seasonality (> 4-fold difference in annual maximum and minimum) was observed in Fuco, which exhibited annual maxima in April and May (monthly mean concentrations ~2 $\mu g\ L^{-1}$) and minima in the fall and winter (mean concentrations < 0.5 $\mu g\ L^{-1}$). On average, TChlb concentrations were relatively high from December to June (maximum of ~0.15 $\mu g\ L^{-1}$ in March), and low from July to October (minimum of ~0.07 $\mu g\ L^{-1}$ in September). Similarly, Hex concentrations were high in winter and spring (~0.2 $\mu g\ L^{-1}$ maximum), and lowest in September (~0.1 $\mu g\ L^{-1}$). Seasonality in Zea was less pronounced, though relatively high concentrations were observed from June to October (annual maximum ~0.06 $\mu g\ L^{-1}$, annual minimum ~0.04 $\mu g\ L^{-1}$). Finally, any potential seasonality in Perid concentrations was not resolvable due to a combination of the predominantly decadal variations observed here (Figure 5C), the bio-optical model's poor performance in reconstructing smaller-scale Perid variations (Figure 4; Table 2), and the coarse sampling resolution available in this data set (see Section 4.2 below for further discussion).

We investigated cross-SBC variability in each PGs annual cycle by quantifying monthly climatologies for each pigment at each PnB station. Past observations of regional advection patterns and satellite sea-surface temperature and chlorophyll *a*

concentrations suggest that the southern PnB stations (PnB station 7 is the southern-most) are generally associated with relatively cool, recently upwelled waters, while the northern PnB stations (e.g., PnB station 1) are associated with warmer surface waters and lower chlorophyll concentrations on average (Harms and Winant, 1998; Henderikx Freitas et al., 2017). These cross-SBC differences are driven by a combination of the upwelling shadow downwind of the coastal Santa Ynez mountains, the physical concentration of phytoplankton by the persistent, convergent eddy in the SBC, and differences in the relative advection of southern CCS and SCBight waters (Harms and Winant, 1998; Simons et al., 2015; Henderikx Freitas et al., 2017). Figure 7 shows the mean +/- 95% confidence intervals in the difference of monthly mean pigment concentrations between stations 1 and 4 (northern vs. central SBC), 1 and 7 (northern vs. southern SBC), and 4 and 7 (central vs. southern SBC).

Amongst the 5 PGs, the largest spatial variations in seasonality were observed for the diatom biomarker pigment, Fuco (Figure 7). Fuco concentrations at station 7 were significantly higher than those at station 1 from June through October, with monthly mean differences of nearly 2 $\mu g\ L^{-1}$ at times (Figure 7H). Fuco concentrations were also significantly higher (> 1 $\mu g\ L^{-1}$ in magnitude) at station 7 relative to station 4 in July, and on average were higher, though not always significantly higher, from April through December (Figure 7I). The opposite pattern was observed in the spatial variations in the annual Zea cycle (Figure 7M-O). Higher Zea concentrations were observed at station 1 and 4 relative to station 7. These differences were most pronounced and often statistically significant from June to December, while differences in monthly mean Zea concentrations between stations 1 and 4 were generally small. Spatial variations in

monthly mean concentrations of TChlb and Hex were relatively small and almost never statistically significant. However, the monthly mean concentrations of both of these pigments were typically lower at station 1 than at stations 4 and 7, except for during the winter and early spring (Figure 7A-F). Dinoflagellates exhibited the opposite pattern, with higher monthly mean Perid concentrations typically found at station 1 relative to stations 4 and 7. However, these differences were never statistically significant due likely to the large amount of variability in Perid on interannual time scales (Figure 7J-L).

Climate forcings and interannual to multi-decadal PG variations

Several notable shifts in the phases of the three major modes of North Pacific climate variability, the ENSO, NPGO, and PDO, occurred during the study period and were linked to interannual variability in some PGs (Figures 8 and 9). Due to their association with a suppression of upwelling winds and water column mixing, extreme El Niño events are expected to result in anomalously low concentrations of the upwelling-responsive PGs (Bograd and Lynn, 2001; Chavez et al., 2002; Venrick, 2012). Such events occurred in 1997-1998 and again in 2015-2016. Mild El Niño events, associated with less severe oceanographic impacts, were also observed in the late fall and early winter of 2002 to 2003, 2004 to 2005, 2006 to 2007, and 2009 to 2010 (Figure 8A-B). Conversely, notable La Niña events, which are expected to enhance upwelling and result in positive anomalies for most PGs, occurred from 1999 to early 2001, 2007 to 2009, and 2010 to 2011.

Relative to the ENSO, both the NPGO and PDO vary on longer time scales. The oceanographic impacts of the NPGO and PDO are less well studied than those of the

ENSO, but generally, the PDO is thought to alter the timing and reduce the amplitude of seasonal upwelling and to drive a relaxation of the California Current in the northern CCS (Mantua and Hare, 2002; Di Lorenzo et al., 2013), while the NPGO has the same impact on the southern CCS and SCBight (Di Lorenzo et al., 2008; Di Lorenzo et al., 2013). The majority of the study period overlapped with the cold phases of both the NPGO and PDO (Figure 8C-D). However, the NPGO briefly shifted to its warm phase from 2005 to early 2007, and again returned to a warm phase for the last five years of the record (Figure 8C). Over the course of the study period, the longest sustained warm phase of the PDO was observed from 2015 to 2017. The PDO was also briefly in its warm phase at the start of the record, and apart from a neutral phase from 2003 to 2006, remained in its cold phase throughout the remainder of the study period (Figure 8D). Notably, both HPLC and bio-optically modeled pigment observations partially overlapped with both cold and warm phases of all three climate oscillations considered here (Supp. Figure S1). While the model training data were biased toward some phases of these climate oscillations (particularly cold phase of the PDO), the model validation results (Figure 4, Supp. Figures S6 and S7) indicate that this bias does not impact model fidelity.

Interannual variations in the seasonal cycles of each PG and in TChla and CHL were investigated by computing anomalies in the concentrations of each biomarker pigment relative to its monthly climatology (Figure 9). After subtraction of each pigment's monthly climatology, anomalies were normalized to monthly mean pigment concentrations so that the anomalies shown in Figure 9 represent a unitless fold-change relative to the monthly climatologies shown in Figure 6.

In general, low anomalies in all five biomarker pigment concentrations were observed for the first five years of the time series (Figure 9). While this period of the time series considers pigment concentrations modeled using the Perkin-Elmer spectrophotometer (see Sections 2.4 and 2.6 above), anomalies in fluorometrically determined CHL concentrations mirror those in modeled TChla concentrations, suggesting these patterns are valid (Figure 9A). A notable commonality amongst the anomaly time series of all PGs except picophytoplankton was the persistence of negative anomalies from 1997-1998 and from 2014-2015. These observations coincided with the two largest El Niño events sampled during this time series (Figure 8), and the latter event was preceded by the extraordinary marine heat wave known as the "warm blob" (Bograd and Lynn, 2001; Bond et al., 2015; Jacox et al., 2016). However, some of the largest positive anomalies in the time series of all 5 biomarker pigments were observed in 2016 while the second extreme El Niño persisted (Figures 8 and 9). More generally, high correlations amongst the anomaly time series of the prymnesiophyte, chlorophyte, and picophytoplankton biomarker pigments were found (Table 4). Conversely, the largest anomalies in diatom and dinoflagellate biomarker concentrations generally did not co-occur. Further, the diatom and dinoflagellate anomaly time series were only weakly correlated with one another and with the other PGs (Table 4).

The largest positive diatom biomarker pigment anomalies were observed in 2008, 2010, 2011, 2016, and 2017 (Figure 9B). Relatively large La Niña events were observed in conjunction with the cold phases of the NPGO and PDO in 2008, 2010, and 2011, which may partially account for these anomalous events (Figure 8). However, a similar alignment of the ENSO, PDO, and NPGO also occurred from 1999-2001 with no

concurrent observations of anomalous diatom blooms (Figures 8 and 9B), suggesting that a complex combination of local and climate forcings are responsible for driving anomalous diatom blooms in the SBC. The highest dinoflagellate anomalies were observed in 2003, 2006, and 2018 (Figure 9C). Taking into account the lack of a resolvable annual cycle in dinoflagellates (Figure 6C), this indicates a roughly decadal pattern in dinoflagellate biomarker pigment concentrations corroborated by previous studies in the SCBight and CCS (Gregorio and Pieper, 2000; Smayda and Trainer, 2010; Fischer et al., 2020). Notably, positive dinoflagellate anomalies were observed during most cruises from 2003 to late 2006, in 2010, and from late 2017 to early 2018; these time periods mostly co-occurred with a negative NPGO index (Figure 8C). The largest positive anomalies in prymnesiophyte pigment concentrations occurred in 2003 and 2012, while the largest chlorophyte pigment anomalies were observed in 2002-2003, 2007, 2009, and 2016 (Figure 9D and 9E). Finally, the largest anomalies in picophytoplankton pigment concentrations were observed from 2007-2009 and in 2016 (Figure 9F). The prymnesiophyte, chlorophyte, and picophytoplankton biomarker pigment anomalies did not demonstrate noticeable event-scale responses to climate forcings other than the previously described suppression of prymnesiophyte and chlorophyte biomarker pigment concentrations during El Niño events (Figures 8 and 9).

We assessed long-term trends in total chlorophyll *a* (Figure 9A) and biomarker pigment concentration anomaly time series (Figure 9B-F), and for monthly anomalies computed separately for each PnB station, using the modified Mann-Kendall trend test for autocorrelated time series outlined in Hamed and Rao (1998) (Supp. Table S2). With the exception of decreasing long-term trends in Hex at PnB stations 2, 3, and 6, no

statistically significant ($p < 0.05$) trends were found for any of the pigments considered (Supp. Table S2). The spatial incoherence of the significant long-term trends in Hex makes them difficult to interpret and suggests they may be due to stochastic variability rather than any oceanographic or climate forcing considered here.

Associations of PGs with oceanographic forcings

We performed empirical orthogonal function (EOF) analysis to determine the dominant modes of association amongst the five biomarker pigments and other oceanographic parameters (Temp, Sal, NO_3, PO_4, and $Si(OH)_4$, representing temperature, salinity, nitrate, phosphate, and silicate, respectively; Figure 10). EOF analysis as applied here is synonymous with principal components analysis and decomposes the data set into a series of orthogonal modes. Each mode is characterized by a set of loadings, or weights, describing the contribution of each variable to the mode (Figure 10A-D), and an amplitude function describing the variability of that mode through time (Figures 10E-H and 11) and, though not considered here, space (Thomson and Emery, 2014). Each mode explains a known fraction of the variance in the original data set, with the first mode accounting for the largest proportion of the total variance and higher order modes accounting for sequentially less variance. For the EOF analysis shown here, all variables were averaged by cruise (spatially) and, where multiple cruises occurred in the same month, by month, to create a monthly time series of each variable as above. Monthly time series of each variable were then standardized to zero mean and unit variance prior to computing EOFs. The first four EOF modes cumulatively explained 82% of the variance in the data set, partitioned across modes 1, 2, 3, and 4 as follows: 40.6%, 19.4%, 11.5%,

10.5%. Modes 5 and 6 explained 6.5 and 5.3% of the variance in the data set, respectively, and all higher order modes explained < 3% of the variance. Modes 5 and 6 are thus not considered here.

The results of the EOF analysis demonstrate the importance of seasonal upwelling responses in driving the variations of the five PGs. The loadings of EOF Mode 1 were positive in the 3 macronutrient concentrations, salinity, TChlb, and Hex, negative in temperature, and only weakly positive for diatom, dinoflagellate, and picophytoplankton pigment concentrations (Figure 10A). This demonstrates a relatively strong covariance amongst chlorophyte and prymnesiophyte pigment concentrations and cold, saline, nutrient-rich waters, indicating rapid positive responses of these PGs to recent upwelling. The mean seasonal cycle of EOF Mode 1 amplitudes confirmed that this mode represents an early upwelling oceanographic state, with the highest monthly mean values observed from March through May and annual minima observed in the late summer and early fall (Figure 10E).

The loadings of EOF Mode 2 were positive in temperature and in all PG biomarker pigment concentrations except for diatoms, and negative in diatom pigment concentrations, salinity, nitrate and phosphate (Figure 10B). This loading pattern indicates contrasting ecosystem states with negative amplitudes corresponding to an upwelling-driven diatom bloom, and positive amplitudes indicating a stratified water column favoring a mixed assemblage dominated by pico- and nano-phytoplankton. Inspection of the mean seasonal cycle of Mode 2 amplitudes again confirmed this interpretation, with annual minima (indicating a diatom bloom) observed in April and May and maxima observed in October and November (Figure 10F). Taken together, EOF

Modes 1 and 2 demonstrate that ~60% of the variance in the combined PG and oceanographic data set is explained by the progressive response of the environment and phytoplankton community to upwelling.

EOF Mode 3 also exhibited some seasonality, but did not appear strongly linked to upwelling dynamics. The loadings of EOF Mode 3 were strongly positive for Perid and to a lesser extent Fuco, while negative loadings were found for both temperature and salinity (Figure 10C). The loadings for the other variables were small indicating relatively weak associations of this mode with the other PGs, as well as with macronutrient concentrations. The mean seasonal cycle of EOF Mode 3 showed positive monthly mean amplitudes during the winter and early spring and negative amplitudes from spring through fall (Figure 10G). This observation in conjunction with the covariance of temperature and salinity loadings in opposition to dinoflagellate pigment concentrations suggests that this mode is associated with winter-time precipitation and freshwater discharge events. Such events are thought to provide favorable conditions for inner-shelf dinoflagellate blooms across the broader California coast (Gregorio and Pieper, 2000; Fischer et al., 2020).

EOF Mode 4 showed positive loadings for diatom and picophytoplankton biomarker pigment concentrations and salinity, in opposition to weak negative loadings in the three macronutrient concentrations (Figure 10D). The mean annual cycle in EOF Mode 4 amplitudes was positive from May to August with an annual maximum in June, and negative throughout late fall, winter, and early spring (Figure 10H). Interpretation of this mode is complicated by the unexpected covariance of diatom and picophytoplankton pigment concentrations with high salinities, but not the other oceanographic properties,

and the potential for the orthogonality constraint of EOF analysis to obscure true oceanographic signals with noise in higher order modes. However, the combination of the observed loading pattern and mean seasonal cycle suggests that this mode may represent a transitional state from a senescing spring diatom bloom to a stratified, picophytoplankton-dominated system. Such a transitional state may be driven by unusually late seasonal upwelling, enhanced entrainment by the persistent cyclonic eddy in the SBC, or some combination of these and other forcings. Inspection of the amplitude time series shows that the highest positive amplitudes of this mode often do not co-occur with the largest negative amplitudes of EOF Mode 2 (which indicates a well-developed diatom bloom), but do co-occur with some of the most anomalous Fuco concentrations observed on PnB (Figures 9B and 11D). Examples of this pattern are most prominent in 2002, 2010, and 2016. Taken together, these results suggest this mode represents a unique diatom bloom state associated with potentially different forcings than the diatom bloom state depicted in EOF Mode 2.

Impacts of climate forcings on PGs and oceanographic modes

Previous work has demonstrated significant impacts of the El Niño Southern Oscillation (ENSO) and the two major modes of Pacific decadal climate variability, the Pacific Decadal Oscillation (PDO) and the North Pacific Gyre Oscillation (NPGO), on coastal California oceanography and marine ecosystems (Mantua et al., 1997; Bograd and Lynn, 2001; Chavez et al., 2002; Di Lorenzo et al., 2008; Venrick, 2012; Di Lorenzo et al., 2013; Jacox et al., 2016; Fischer et al., 2020). Less is known about the impacts of these climate oscillations on PG dynamics in the SBC (Anderson et al., 2008; Venrick,

2012; Barth et al., 2020; Fischer et al., 2020), though the qualitative associations with the PG biomarker pigment concentrations described above indicate non-negligible impacts (Figures 8 and 9). To further investigate the roles of these climate oscillations in driving event-scale PG variations in the SBC, conditional averages of the seasonal anomalies of each biomarker pigment and EOF mode amplitude function (Figure 12) were computed for the 15% largest positive and negative values of each of the four climate indices (the Southern Oscillation Index, or SOI, and Multivariate ENSO Index, or MEI, provide two independent indices of ENSO) overlapping with the biomarker pigment (Figure 12A-E) and pigment-oceanographic time series used to compute EOFs (Figure 12F-I). Statistically significant differences between conditionally averaged pigment concentrations or EOF mode amplitudes were assessed using two-sample t-tests.

The impacts of ENSO events on the dynamics of specific PGs and PG-oceanographic EOF modes were evident in the conditional averaging of PG biomarker pigment and EOF mode amplitude anomalies (Figure 12). As expected, diatom pigment concentrations showed positive anomalies during La Niña events and negative anomalies coincident with El Niño events. These impacts were consistent in both pattern and magnitude for both the SOI and MEI (Figure 12C), and corroborated by significant correlations between Fuco and both the SOI and MEI (Table 5). The impacts of the ENSO on the other biomarker pigments was less clear. Conditionally averaged prymnesiophyte pigment concentration anomalies showed contrasting patterns across the two ENSO indices and neither difference was statistically significant, indicating no observable impact (Figure 12B). The patterns of conditionally averaged dinoflagellate pigment concentration anomalies across the SOI and MEI were similar, with La Niña

events favoring higher concentrations, though the differences were not statistically significant (Figure 12D). Significantly higher (lower) anomalies were observed during La Niña events for the chlorophyte (picophytoplankton) biomarker pigment concentrations when considering the SOI (MEI), but no observable effects on these biomarker pigments were found according to the MEI (SOI) (Figure 12A and 12E). No significant correlations were found between TChlb or Zea with either the MEI or SOI (Table 5).

The decadal modes of North Pacific Climate variability, the PDO and NPGO, also had variable impacts on the interannual dynamics of each PG. Significant differences were observed in conditionally averaged dinoflagellate pigment concentration anomalies for both the NPGO and PDO (Figure 12D). The magnitude of differences was greater for the NPGO and Perid was significantly correlated with the NPGO (Table 5) but was not significantly correlated with the PDO, SOI, or MEI. Interestingly, high dinoflagellate anomalies were favored by the warm phase of the NPGO, but suppressed by the warm phase of the PDO. Conversely, the cold phase of the PDO significantly favored anomalously high diatom pigment concentrations, while the NPGO did not have an obvious impact on interannual Fuco variations (Figure 12C, Table 5). We observed no significant differences in conditionally averaged prymnesiophyte, chlorophyte, or picophytoplankton pigment concentration anomalies, and no significant correlations between these pigment concentration anomalies and the NPGO or PDO (Figure 12A, 12B, 12E, Table 5). However, higher anomalies in prymnesiophyte and chlorophyte pigment concentrations were observed during the cold phase of the PDO and NPGO relative to their respective warm phases.

Conditional averaging of the four EOF mode amplitude anomalies according to the 15% largest positive and negative values of the four climate oscillations that overlapped with the relevant pigment and oceanographic observations largely corroborated the results of the conditional averaging of the anomalies of each biomarker pigment (Figure 12F-I). EOF Mode 1, corresponding to an early-upwelling state with cold temperatures and high TChlb, Hex, and macronutrient concentrations, was significantly impacted by all three climate oscillations (Figure 12F, Table 5). The cold phase of all 3 climate oscillations favored anomalously strong upwelling and high chlorophyte and prymnesiophyte pigment concentrations. Similarly, the cold phases of the PDO, NPGO, and ENSO all favored negative amplitudes of EOF Mode 2, a proxy for diatom blooms, although the conditional averages were only significantly different for the MEI (Figure 12G). Correlations between EOF Mode 2 anomalies and the MEI, NPGO, and PDO were also significant (Table 5). Conditionally averaged values of EOF Mode 3 amplitude anomalies, interpreted as a dinoflagellate bloom mode associated with winter-time discharge events, showed significant impacts of both the NPGO and PDO, with the warm (cold) phase of the NPGO (PDO) favoring anomalous dinoflagellate blooms (Figure 12H). Both indices of the ENSO suggested an enhancement of dinoflagellate blooms during La Niña events. These results were corroborated in part by significant correlations between EOF Mode 3 anomalies and the SOI and NPGO (Table 5). Finally, the transition state from a well-developed diatom bloom to a picophytoplankton dominated assemblage associated with high salinity surface waters, indicated by positive amplitudes of EOF Mode 4, was favored by cold phases of the ENSO, PDO, and NPGO, although significant differences in the conditionally averaged amplitudes of Mode 4,

along with a significant correlation coefficient, were only observed for the NPGO (Figure 12I, Table 5).

Discussion

Summary of results

We quantified seasonal to multi-decadal PG dynamics in the SBC based on an approximately monthly time series of HPLC and bio-optically modeled biomarker pigment concentrations spanning more than 20 years. The dominant SBC PGs resolvable from these HPLC pigment concentration data were identified using hierarchical cluster analysis and included diatoms, dinoflagellates, chlorophytes, prymnesiophytes, and picophytoplankton (Figure 2). The concentrations of five biomarker pigments, each assumed to represent the pigment biomass of one of the above PGs, were modeled with high fidelity using a previously developed bio-optical modeling approach (Table 2, Figures 3 and 4, Supp. Figures S6 and S7). Seasonal variations were resolvable for all PG biomarker pigments except Perid (representing dinoflagellates). On average, seasonal variations ranged from ~1.5-fold for the picophytoplankton, to ~2-fold for the prymnesiophytes and chlorophytes, to >4-fold for diatoms (Figure 6). The magnitude and patterns of each PG's annual cycle showed significant cross-SBC differences (Figure 7). Relative to monthly mean biomarker pigment concentrations, interannual variations were as high as 2-3-fold for picophytoplankton, prymnesiophytes, and chlorophytes, and occasionally larger than 5-fold for both the diatoms and dinoflagellates (Figure 9). To the extent that PG dynamics were associated with oceanographic forcings, upwelling exerted the strongest control on PG dynamics (Figures 10 and 11). Natural climate oscillations

including the ENSO, PDO, and NPGO exhibited unique associations with each PG at the event scale (Figures 8, 9, 11, 12).

In the following, we discuss the limitations of the present study for assessing long-term PG dynamics. We then explore the role of regional surface advection patterns in driving some of our observations of seasonal to multi-decadal PG dynamics. Finally, we place our observations of the associations of the dominant SBC PGs with oceanographic and climate forcings in the context of broader knowledge of large-scale PG dynamics in the California Current System (CCS), Southern California Bight (SCBight), and more generally in upwelling systems.

Limitations of the present study – what are we missing?

The assessments of seasonal to multi-decadal PG dynamics presented above rely on a synthesis of HPLC and bio-optically modeled phytoplankton pigment concentrations to create a 22-year, approximately monthly record of PG biomarker pigment concentrations. We have shown in Section 2.6 that the methods used to synthesize these two data sets are robust and well-validated with independent data. However, several major limitations remain to be addressed in order to use these data to assess seasonal to multi-decadal PG dynamics. Here, we discuss these limitations and how they may impact our interpretations of the results presented above.

Like all methods for assessing PG dynamics, HPLC pigment analysis has strengths and weaknesses (Lombard et al., 2019). The prominent strengths of the HPLC method are demonstrated in our analysis: rigorously evaluated and standardized analytical procedures (Van Heukelem and Thomas, 2001; Hooker et al., 2010) enable

precise and accurate PG observations with high spatiotemporal coverage; unique absorption signatures of biomarker pigments found in *in situ* and remotely sensed bio-optical properties allow biomarker pigment concentrations to be modeled with high skill, expanding the spatiotemporal coverage of observations (Chase et al., 2017; Catlett and Siegel, 2018); and the PGs resolved by pigment methods span a more holistic range of phytoplankton size classes than possible for many other methods.

However, HPLC (and bio-optically modeled) pigment concentrations also have widely-documented limitations and uncertainties (Higgins et al., 2011; Jeffrey et al., 2011). First, investigators must assume that biomarker pigment concentrations reasonably approximate the biomass of PGs. This assumption is applied explicitly here, and implicitly (often with additional assumptions) in studies employing more complicated pigment chemotaxonomy methods (Mackey et al., 1996; Uitz et al., 2006; Hirata et al., 2011). However, variability in pigment concentrations can arise due to a combination of changes in PG biomass, physiological responses to environmental conditions, and genetic or other sources of intra-PG variability (Higgins et al., 2011; Kramer and Siegel, 2019). In particular, the 1.5- to 2-fold seasonal variations observed in TChlb, Hex, and Zea above (Figure 6) fall within a range that could be explained by physiological variability in pigmentation (see Higgins et al., 2011, and references therein). Further, comparisons of PG dynamics inferred from photoprotective (Zea in the present analysis) and photosynthetic (including TChlb, Hex, Fuco, and Perid) pigments may be susceptible to biased interpretations given that these pigments vary differently in response to changing irradiance (Higgins et al., 2011). However, the cluster analysis (Figure 2) shows that to first order, Zea covaries more strongly with DVChla (a photosynthetic pigment) than

other photoprotective pigments, while photosynthetic biomarker pigments representative of other PGs covary with distinct suites of photoprotective pigments. These results suggest that variability in PG biomass is the first order determinant of the variations in biomarker pigment concentrations observed here.

Another limitation of biomarker pigment assessments is the ambiguity in the representation of a single PG by a particular biomarker pigment (Jeffrey et al., 2011; Kramer and Siegel, 2019). For example, the diatom biomarker pigment Fuco is also found in many other PGs, including the oft-abundant dinoflagellates, prymnesiophytes, and pelagophytes (Jeffrey et al., 2011). Bloom-forming dinoflagellates occasionally obscure the Fuco-diatom relationship in the SBC (Catlett et al., 2020b). This may partially explain certain anomalous Fuco observations in the present analysis, such as the ~2-fold anomaly in the fall of 2003 (Figure 9) and the covariation of Fuco with the dinoflagellate biomarker pigment Perid in EOF Mode 3 (Figure 10). Conversely, Perid is not found in some lineages of photosynthetic dinoflagellates and so is not representative of this entire PG (Jeffrey et al., 2011), which may partially explain the lack of a resolvable annual cycle in our results above (Figure 6). Finally, the underlying genetic, taxonomic, and functional diversity represented by each PG and biomarker pigment is highly variable. For example, the prymnesiophytes include a diverse array of functional groups including calcifiers like *Emiliania huxleyi*, DMS producers like *Phaeocystis sp.*, and mixotrophs like *Prymnesium parvum* (Nygaard and Tobiesen, 1993; Van Boekel and Stefels, 1993; de Vargas et al., 2007). Given their diverse functional roles and ecological niches, each of these prymnesiophyte species may be expected to respond differently to oceanographic and climate forcings. Thus, the lack of clear associations of some PGs

with oceanographic and climate forcings above may be explained in part by intra-PG variability in responses to these forcings.

The other primary limitation of the present study is that the monthly, unevenly sampled time series presented here does not capture short-term PG dynamics and failed to resolve some anomalous events. One known example of this is the unprecedented *E. huxleyi* bloom that occurred in the SBC in the first week of June, 2015 (Matson et al., 2019). After initial detection of the bloom in satellite imagery on May 31, 2015 (17 days after the PnB cruise in May, 2015), Matson et al. (2019) observed *E. huxleyi* cell concentrations on June 4, 2015 that were an order of magnitude greater than had ever been previously observed in the SCBight. Satellite imagery showed the bloom began decaying shortly after June 4 and had largely dissipated by the time PnB observed Hex concentrations similar to climatological mean values on June 18, 2015 (Matson et al., 2019; Figures 5D and 9E). There were likely additional anomalous blooms of specific PGs that were not sampled by PnB over the course of the 20+ year record presented here. The chronic under-sampling of January and February in this time series due to ship availability also leads to greater uncertainty surrounding typical winter-time PG concentrations in the SBC. Nonetheless, the broad seasonal and interannual patterns highlighted in the above analyses are likely robust to the imperfect sampling of the PnB time series and largely corroborate and complement existing observations of large-scale PG dynamics in the CCS and SCBight (see Section 4.4 below).

Finally, the 22-year biomarker pigment time series provides a rare glimpse into the climate forcings of interannual to multi-decadal PG dynamics. However, assessing the roles of the ENSO, PDO, and NPGO in driving interannual to decadal PG dynamics

in the SBC remains difficult given the paucity of climate phase transitions and extreme events observed over the 22-year biomarker pigment record. Only two strong El Niño events were sampled over the course of the time series, both of which coincided with warm phases of the NPGO and PDO (Figure 8). Similarly, both the NPGO and PDO remained in the cold phase for the majority of the 22-year time series, with only two significant, though relatively brief warm events sampled for each climate index. Further, the two warm PDO events coincided with extreme El Niño events, and the latter also coincided with the anomalous "warm blob" event in 2014-15 (Bond et al., 2015). Given these small sample sizes (despite the inclusion of 1393 stations sampled on 238 PnB cruises conducted over 22 years) and the potential for interactions amongst these climate oscillations, robust assessments of the impacts of the ENSO, PDO, and NPGO, as well as their interactions with one another and with anthropogenic climate forcing, on PG dynamics will require substantially longer time series than the 22-year record presented here. Nonetheless, the analyses above (Figures 8, 9, 12) provide an important step towards determining the roles North Pacific climate variability and anthropogenic climate forcing will play in determining SBC PG dynamics in the future.

Potential roles of advection in driving PG dynamics

The SBC's location in the transition zone between the upwelling-impacted, nutrient-rich waters of the CCS and the warmer, more oligotrophic waters of the SCBight, and the prevailing circulation patterns in and around the SBC (Harms and Winant, 1998; Winant et al., 2003; Dong et al., 2009; Brzezinski and Washburn, 2011), suggest that advection of source waters from these adjacent environments may impact observations of

SBC PG dynamics. In general, upwelling winds intensify equatorward flows in the CCS and result in the advection of southern CCS waters into the western entrance of the SBC (Harms and Winant, 1998; Brzezinski and Washburn, 2011). The relaxation of upwelling winds allows for a return flow of SCBight waters poleward along the mainland coast and into the eastern entrance of the SBC (Harms and Winant, 1998; Melton et al., 2009). On longer time scales, the NPGO appears related to variations in the balance of these two flows in the southern CCS and SCBight (Di Lorenzo et al., 2008; Di Lorenzo et al., 2013).

While there is high variability in surface ocean circulation patterns in and around the SBC, direct observations of surface currents and spatial patterns of satellite sea surface temperature and chlorophyll *a* concentrations apparently confirm that the southern and western portions of the SBC tend to be more heavily impacted by CCS waters, while the northern and eastern SBC are more frequently impacted by SCBight waters (Harms and Winant, 1998; Henderikx Freitas et al., 2017). In conjunction with past studies showing more frequent dominance by diatoms (dinoflagellates) in CCS (SCBight) waters (Venrick, 2002; Venrick, 2012; Taylor et al., 2015), our observations of the prominent spatial variations in PG seasonality (Figure 7), as well as the decadal dinoflagellate anomalies associated with the NPGO (Figures 8, 9, 11, and 12), suggest an important role of advection in driving seasonal to multi-decadal PG dynamics in the SBC.

Here we employ a Lagrangian particle tracking model within a high resolution ROMS solution for a 10-year subset (2004-2013) of the PG time series to investigate whether variations in source water origin alters phytoplankton community composition in the SBC (see Section 2.8.2). Particles were tracked backwards in time from 34 release points along the PnB transect on each day of the 10-year time series. Figure 13 shows

examples of particle trajectories projected backwards in time for two different days of the time series where a majority of particles originated from the West origin box (Figure 13A) or the East origin box (Figure 13B). These examples provide a synoptic view of advection patterns over a short-term (~2 week) upwelling-relaxation cycle, and support the assumption that on 10-day advection time scales, the proportion of particles advected from the West (East) origin box provides a reasonable approximation for the relative influence of southern CCS (SCBight) source waters on SBC PG dynamics. In order to align further analyses of the simulated particle trajectories with the approximately monthly sampling of PnB, monthly time series of the proportion of particles derived from each origin box for the PnB transect and for each PnB station were computed from the daily time series and are discussed here (see Section 2.8.2). All results considered here are for 10-day advection times; results from 5- and 15-day advection times are presented in Supporting Figures S9 and S10, and qualitatively agree with those shown here.

First, we used the simulated source water assessments to test the hypothesis that on seasonal time scales, cross-SBC variability in climatological mean diatom, dinoflagellate, and picophytoplankton pigment concentrations (Figure 7) are driven by variations in source water origin. Increased advection of CCS (SCBight) source waters is expected to lead to seasonally elevated diatom (dinoflagellate and picophytoplankton) concentrations. Figure 14A shows the mean annual cycle in the proportion of particles originating from each origin box for the PnB transect. The seasonal cycle in SBC source waters appeared to be tightly coupled to seasonal upwelling. Across the PnB transect, the proportion of particles from the West origin box (Figures 1 and 13), a proxy for the magnitude of advection of CCS sources waters, was highest in March and elevated

(>20%) from February through May (Figure 14A). Conversely, advection of SCBight waters into the SBC as indicated by the proportion of particles originating from the East origin box was lowest in March and April and subsequently increased until reaching an annual maximum in September (Figure 14A), as expected (Harms and Winant, 1998). Relatively few particles reached the PnB transect from the South origin box.

Seasonal cycles in the source waters of the northern- and southern-most portions of the SBC deviated substantially from that observed for the PnB transect as a whole (Figure 14C-D). From March through June, the proportion of particles originating from the CCS at the southern-most release points (closest to PnB station 7) was >10% higher than observed for the transect, and was >5% higher throughout the remainder of the year (Figure 14D). The opposite pattern was observed at the release points closest to PnB station 1 in the northern SBC (Figure 14C). This cumulatively represents a ~20-40% difference in CCS source water advection between PnB stations 1 and 7 from March through June (Figure 14). These differences are associated with 1-2 $\mu g\ L^{-1}$ higher monthly mean diatom biomarker pigment concentrations at PnB station 7 relative to station 1 from March through September, and smaller but significantly different picophytoplankton concentrations at station 7 relative to station 1 (Figure 7). Shorter-term studies have previously documented advection of harmful diatom blooms associated with elevated domoic acid concentrations from the southern CCS into the southwestern SBC during the late summer and fall (Anderson et al., 2009). Our results suggest this phenomenon may be a consistent source of elevated phytoplankton concentrations in the SBC during the late summer and fall. Consistently higher dinoflagellate concentrations in the northern SBC (station 1) relative to the southern SBC (Figure 7) are also linked to

consistently higher advection of SCBight source waters (Figure 14). Overall, these findings support our hypothesis that the relative magnitude of advection of CCS and SCBight source waters into the SBC plays a substantial role in driving spatial variations in SBC PG dynamics on seasonal and interannual time scales.

On interannual to decadal time scales, we hypothesized that NPGO-driven variations in the advection of SCBight source waters into the SBC (Di Lorenzo et al., 2008; Di Lorenzo et al., 2013) provide favorable conditions and/or seed dinoflagellate populations enabling the anomalous decadal dinoflagellate blooms observed above to develop (Figures 5, 8, 9, 11, and 12). We test this hypothesis using the simulated source water determinations for the entire PnB transect by comparing monthly anomalies in the proportion of particles advected from each origin box with the observed NPGO and dinoflagellate biomarker pigment dynamics (Figure 15). Anomalously high advection of SCBight source waters was found almost every month from 2004 to late 2007 (Figure 15A), coupled with a warm phase of the NPGO (Figure 15B) and consistently high dinoflagellate concentrations (Figure 15C). Conversely, from late 2007 through 2013, the cold phase of the NPGO was coupled with only sporadic observations of anomalously high advection of SCBight source waters and dinoflagellate concentrations.

Consideration of the monthly dynamics leading to the dinoflagellate bloom in early 2006 further supports this hypothesis. An anomalously large red tide dominated by the Perid-containing *Lingulodinium polyedrum* (Zapata et al., 2012) was observed over a large extent of the nearshore SCBight from March through September in 2005 (Santoro et al., 2010). Monthly mean Perid concentrations in the SBC rose from 0.22 $\mu g\ L^{-1}$ in August 2005 to ~0.7 $\mu g\ L^{-1}$ in October and November 2005, coinciding with anomalously

high advection of SCBight source waters from July to October 2005 (Figure 15C). This pattern was interrupted by a highly anomalous intrusion of CCS source waters from November 2005 to January 2006 (Figure 15A), interpreted here as an introduction of nutrient-rich waters to the SBC that enabled growth and accumulation of dinoflagellates (monthly mean Perid concentration of 2.39 $\mu g\ L^{-1}$ in January, 2006; see Figure 15C). While other factors must align to allow for the accumulation of dinoflagellates in SBC surface waters, these results highlight the importance of advection in driving the anomalous decadal dinoflagellate blooms observed above (Figure 5).

Altogether, the ROMS backwards particle tracking simulations provide strong evidence that our Eulerian observations of seasonal to multi-decadal PG dynamics in the SBC are impacted by variability in the advection of different source waters into the SBC. Seasonally elevated diatom biomarker pigment concentrations in the southwestern SBC were associated with seasonally elevated advection of CCS source waters, while decadal dinoflagellate blooms were associated with anomalously high advection of SCBight source waters linked to the warm phase of the NPGO. However, it remains unknown whether these contrasting source waters harbor "seed" populations of PGs that are primed for or in the midst of blooming, or if local SBC PG populations are favored by the oceanographic properties of the source waters. Further targeted studies focusing on synoptic perspectives of PG bloom events in addition to genetic studies of SBC PG populations relative to those found in the CCS and SCBight may resolve this question. Regardless, seasonal to interannual variability in source water origins should be accounted for in studies of long-term PG dynamics, particularly in oceanographic transition zones like the SBC.

Oceanographic and climate forcing of PG dynamics in the SBC

Wind-driven upwelling has long been recognized as the dominant forcing of seasonal to interannual variations in phytoplankton biomass, productivity, and community composition in the SBC, CCS, and SCBight (Goodman et al., 1984; Venrick, 2002; Anderson et al., 2008; Barth et al., 2020; Fischer et al., 2020). The focus of most studies to date has been on the seasonal "succession" (though this is not equivalent to succession as traditionally defined by ecologists; see Barber and Hiscock, 2006) of the phytoplankton community from a diatom-dominated community during periods of significant spring upwelling to a dinoflagellate-dominated community as the water column becomes more stratified following the relaxation of upwelling in summer and fall (Margalef, 1978; Goodman et al., 1984; Anderson et al., 2008; Barth et al., 2020; Fischer et al., 2020). Due to limitations of methods relying on visual identification of PGs, the seasonal dynamics of nano- and pico-phytoplankton groups are often not considered, although long-term epifluorescence microscopy observations have documented some seasonal and interannual variations in pico- and nano-phytoplankton groups (Taylor et al., 2015; Caron et al., 2017). Here we discuss the complimentary view of the responses of the phytoplankton community, particularly pico- and nano-phytoplankton groups, to seasonal upwelling and climate forcings provided by the biomarker pigment time series presented above in the context of past studies reliant on microscopic PG observations in upwelling systems.

The EOF analysis of biomarker pigments and oceanographic observations above (Figures 10 and 11) shows the progressive responses of different PGs to seasonal

upwelling. While the diatoms tend to reach the highest overall cell densities (Anderson et al., 2006; Venrick, 2012; Taylor et al., 2015; Caron et al., 2017) and pigment biomass (Anderson et al., 2008; Figure 6) in response to upwelling, the loading pattern and monthly mean amplitudes of EOF Mode 1 show that the typical "first responders" to seasonal upwelling in the SBC are the chlorophytes and prymnesiophytes (Figures 6 and 10). On average, the annual peak in diatom pigment biomass occurs in April or May, after the initial peak in chlorophyte and prymnesiophyte pigment biomass in March (Figures 6 and 10). These results are consistent with previous observations in the SBC and SCBight showing high winter-time abundances of nano-phytoplankton and annual maxima in prymnesiophyte abundances in the early spring (Taylor et al., 2015; Caron et al., 2017). Similarly, some pico- and nano-phytoplankton respond positively to elevated nutrient concentrations in the broader CCS as well as in the equatorial Pacific upwelling zone (Barber and Hiscock, 2006; Taylor and Landry, 2018). Reduced top-down regulation of diatoms relative to smaller-sized PGs likely explains the tendency for diatoms to accumulate more biomass than smaller PGs in response to favorable growth conditions (see Taylor and Landry, 2018 for a detailed discussion), though further study is needed to confirm this hypothesis in the SBC. Regardless, these observations suggest that assumptions of a neutral or negative response of all pico- and/or nano-phytoplankton to elevated nutrient concentrations often employed in marine ecosystem models should be revisited, as suggested previously (Barber and Hiscock, 2006; Taylor and Landry, 2018).

Interestingly, we did not find an obvious pattern of phytoplankton community "succession" from a diatom bloom in spring/summer to a period of elevated

dinoflagellate concentrations in summer/fall as might be predicted in some interpretations of Margalef's mandala (Margalef, 1978; Figures 6 and 10). The large multi-decadal variations in Perid concentrations combined with the poor performance of the bio-optical model when predicting low Perid concentrations (Figure 4; Supp. Figures S6 and S7) may have obscured underlying seasonal dinoflagellate variations. However, the mean annual cycles of Fuco and Perid at PnB station 1 on the mainland shelf showed signs of the dynamics predicted by Margalef's mandala (Supp. Figure S11), and publicly available microscopy observations at the nearby Stearns Wharf, Santa Barbara, CA often show a seasonal increase in cell abundances of some dinoflagellates beginning in the late spring and early summer and extending into the early fall (Supp. Figure S12). These findings support previous suggestions of a decoupling of PG dynamics on the inner continental shelf (water depths < ~30 to 40 m) from those observed further offshore in the SBC, SCBight, and central CCS (Lucas et al., 2011; Goodman et al., 2012; Schulien et al., 2017), and more broadly demonstrate the importance of pairing near-shore marine ecosystem monitoring programs (e.g., SCCOOS) with offshore observations. In agreement with previous studies (Gregorio and Pieper, 2000; Fischer et al., 2020), the covariance of temperature and salinity loadings in opposition to the Perid loading in EOF Mode 3 (Figures 10 and 11) suggest that sporadic winter-time precipitation and freshwater discharge events are likely a more prominent forcing of dinoflagellate blooms in the broader SBC region than seasonal relaxations of upwelling winds.

Interannual variations in the oceanographic manifestations of seasonal upwelling are largely dictated by climate forcings, most notably the ENSO (Bograd and Lynn, 2001; Venrick, 2012; Jacox et al., 2016), PDO (Mantua et al., 1997; Jacox et al., 2014),

and NPGO (Di Lorenzo et al., 2008; Di Lorenzo et al., 2013; Jacox et al., 2014). El Niño events drive an anomalously stratified water column and deepening of the nutricline in the SCBight and CCS, which generally leads to anomalously low phytoplankton biomass (Bograd and Lynn, 2001; Venrick, 2012). The impacts of the ENSO on SBC PG dynamics are demonstrated above (Figures 8, 9, 11, 12). The two strongest El Niño events (1997-98 and 2015-16) over our 22 years of observations were accompanied by anomalously low pigment biomass for 4 of the 5 PGs investigated (all except picophytoplankton; Figures 8 and 9). The conditional averaging of EOF amplitudes by the two indices of the ENSO confirm that La Niña events favor enhanced upwelling and the associated responses of the chlorophytes, prymnesiophytes, and diatoms in EOF Modes 1, 2 and 4, while El Niño events favor higher picophytoplankton concentrations (Figure 12). Dinoflagellate concentrations as indicated by Perid and EOF Mode 3 also appear higher during La Niña events despite a lack of clear associations with the oceanographic signatures of upwelling (Figure 12).

Although a 22-year time series only offers a limited view of decadal processes, our observations provide a glimpse into low-frequency PG variations governed by the NPGO and PDO. The dominant decadal pattern observed in the PG data set was the anomalous dinoflagellate blooms associated with the warm phase of the NPGO and the cold phase of the PDO (Figures 8, 9, 11, and 12). Anomalously high dinoflagellate abundances have been recently observed in association with the warm phase of the NPGO on the inner shelf of Central California (Barth et al., 2020; Fischer et al., 2020), suggesting this association may hold for a significant portion of the CCS and SCBight. In conjunction with the remote forcing of the NPGO, these studies have proposed a

combination of increased freshwater discharge events (also corroborated by our analysis; Figures 10 and 12) and increased water column stratification in driving these dinoflagellate anomalies (Barth et al., 2020; Fischer et al., 2020). Figures 14 and 15 above suggest that enhanced advection of SCBight source waters plays an important role in driving these blooms in the SBC, though it is unclear if this phenomenon would extend north of Point Conception.

The NPGO and PDO are also expected to impact PG responses to seasonal upwelling (Mantua et al., 1997; Di Lorenzo et al., 2008; Chenillat et al., 2012; Di Lorenzo et al., 2013). The cold phases of the NPGO and PDO apparently favored the upwelling responsive PGs identified above (diatoms, prymnesiophytes, and chlorophytes; see Figure 12). Interestingly, EOF Modes 2 and 4 resolved two independent (by definition of the EOF analysis) diatom bloom states. While both were favored by the cold phases of the ENSO, NPGO, and PDO, EOF Mode 2 showed an annual maximum in May and was significantly impacted by the ENSO (Figure 12; $r = 0.19$, $p = 0.008$), but EOF Mode 4 showed an annual maximum in June and was significantly impacted by the NPGO (Figure 12; $r = 0.16$, $p = 0.03$). The mechanisms driving these differences are not clear and require further exploration, though both the ENSO and NPGO are thought to impact the timing of seasonal upwelling in the broader CCS and this may partially explain these results (Bograd et al., 2009; Chenillat et al., 2012). These results suggest that extreme NPGO and ENSO events, along with the associated impacts on oceanographic and other forcings, may lead to unique realizations of seasonal diatom bloom dynamics.

Taken together, our results reveal the seasonal to multi-decadal oceanographic and climate forcings of the dominant PGs in and around the SBC. In addition to diatoms, several smaller-sized PGs accumulate pigment biomass in response to seasonal upwelling in the SBC. All PGs that exhibit seasonal variability associated with upwelling are impacted by interannual variations in seasonal upwelling linked to forcing by the ENSO, NPGO, and PDO. Seasonal variability in source water advection drives pronounced cross-SBC variability in annual PG cycles, particularly in the magnitude of diatom blooms, while decadal dinoflagellate blooms in the SBC are linked to the NPGO, freshwater discharge, and multi-decadal variations in regional advection patterns. Future research is required to determine the roles of top-down forcings in shaping the dynamics of phytoplankton communities in the SBC, as well as to clarify the underlying mechanisms linking decadal dinoflagellate blooms to anomalous regional advection patterns.

Acknowledgements

We thank the past and present members of the Plumes and Blooms team for their many years of dedicated sampling and analysis efforts, and acknowledge helpful conversations with Debora Iglesias-Rodriguez, Libe Washburn, Chris Gotschalk, Mark Brzezinski, and Sasha Kramer. We also thank three anonymous peer reviewers for helpful feedback on the manuscript. Plumes and Blooms is funded by the National Aeronautics and Space Administration (NASA; NNX11AL94G), and ship time for the data presented here was provided by NASA and the National Oceanic and Atmospheric Administration (NOAA) Channel Islands National Marine Sanctuary. Additional support

of this work and of DC was provided by the NASA Biodiversity and Ecological Forecasting program (Grant NNX14AR62A), the Bureau of Ocean and Energy Management Ecosystem Studies program (BOEM award MC15AC00006) and NOAA in support of the Santa Barbara Channel Marine Biodiversity Observation Network, the NASA PACE Science Team (Grant 80NSSC20M0226), and the NSF Santa Barbara Coastal Long Term Ecological Research site (Grant OCE 1232779). DC was also funded in part by a NASA Earth and Space Science Fellowship (Grant NNX16AO44HS02).

References

Allen WE. 1942. Occurrences of "red water" near San Diego. *Science* **96**(2499): 471–471.

Anderson CR, Brzezinski MA, Washburn L, Kudela R. 2006. Circulation and environmental conditions during a toxigenic Pseudo-nitzschia australis bloom in the Santa Barbara Channel, California. *Marine Ecology Progress Series* **327**: 119–133.

Anderson CR, Siegel DA, Brzezinski MA, Guillocheau N. 2008. Controls on temporal patterns in phytoplankton community structure in the Santa Barbara Channel, California. *Journal of Geophysical Research: Oceans* **113**(C4).

Anderson, CR, Siegel, DA, Kudela, RM, Brzezinski, MA. 2009. Empirical models of toxigenic Pseudo-nitzschia blooms: potential use as a remote detection tool in the Santa Barbara Channel. *Harmful Algae*, **8**(3): 478-492.

Barber R, Hiscock M. 2006. A rising tide lifts all phytoplankton: Growth response of other phytoplankton taxa in diatom-dominated blooms. *Global Biogeochemical Cycles* **20**(4). Wiley Online Library.

Barrón RK, Siegel DA, Guillocheau N. 2014. Evaluating the importance of phytoplankton community structure to the optical properties of the Santa Barbara Channel, California. *Limnology and oceanography* **59**(3): 927–946. Wiley Online Library.

Barth A, Walter RK, Robbins I, Pasulka A. 2020. Seasonal and interannual variability of phytoplankton abundance and community composition on the Central Coast of California. *Marine Ecology Progress Series* **637**: 29–43.

Beers J. 1986. Organisms and the food web. *Plankton dynamics of the southern California Bight* **15**: 84–175. Wiley Online Library.

Behrenfeld MJ, Boss ES. 2018. Student's tutorial on bloom hypotheses in the context of phytoplankton annual cycles. *Global change biology* **24**(1): 55–77. Wiley Online Library.

Bialonski S, Caron DA, Schloen J, Feudel U, Kantz H, Moorthi SD. 2016. Phytoplankton dynamics in the Southern California Bight indicate a complex mixture of transport and biology. *Journal of Plankton Research* **38**(4): 1077–1091.

Bograd SJ, Lynn RJ. 2001. Physical-biological coupling in the California Current during the 1997–99 El Niño-La Niña cycle. *Geophysical Research Letters* **28**(2): 275–278. Wiley Online Library.

Bograd SJ, Schroeder I, Sarkar N, Qiu X, Sydeman WJ, Schwing FB. 2009. Phenology of coastal upwelling in the California Current. *Geophysical Research Letters* **36**(1).

Bond NA, Cronin MF, Freeland H, Mantua N. 2015. Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophysical Research Letters* **42**(9): 3414–3420.

Brzezinski MA, Washburn L. 2011. Phytoplankton primary productivity in the Santa Barbara Channel: Effects of wind-driven upwelling and mesoscale eddies. *Journal of Geophysical Research: Oceans* **116**(C12).

Caron DA, Connell PE, Schaffner RA, Schnetzer A, Fuhrman JA, Countway PD, Kim DY. 2017. Planktonic food web structure at a coastal time-series site: I. Partitioning of microbial abundances and carbon biomass. *Deep Sea Research Part I: Oceanographic Research Papers* **121**: 14–29. Elsevier.

Carr SD, Capet XJ, McWilliams JC, Pennington JT, Chavez FP. 2008. The influence of diel vertical migration on zooplankton transport and recruitment in an upwelling region: Estimates from a coupled behavioral-physical model. *Fisheries Oceanography* **17**(1): 1–15. Wiley Online Library.

Catlett D, Siegel DA, Guillocheau N. 2020a. Plumes and Blooms: Curated oceanographic and phytoplankton pigment observations ver 1. Environmental Data Initiative. https://doi.org/10.6073/pasta/f88ee1dc32b8785fe6ce57d80722e78c.

Catlett D, Siegel DA, Carlson CA, Matson PG, Wear EK, Iglesias-Rodriguez MD. 2020b. Integrating high-throughput sequencing observations into remotely sensible phytoplankton functional type determinations. Ocean Sciences Meeting 2020. AGU.

Catlett D, Siegel DA. 2018. Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment. *Journal of Geophysical Research: Oceans* **123**(1): 246–264.

Chase A, Boss E, Cetinić I, Slade W. 2017. Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra: toward a global algorithm. *Journal of Geophysical Research: Oceans* **122**(12): 9725–9743. Wiley Online Library.

Chavez FP, Pennington JT, Castro CG, Ryan JP, Michisaki RP, Schlining B, Walz P, Buck KR, McFadyen A, Collins CA. 2002. Biological and chemical consequences of the 1997–1998 El Niño in central California waters. *Progress in Oceanography* **54**(1–4): 205–232.

Chenillat F, Rivière P, Capet X, Di Lorenzo E, Blanke B. 2012. North Pacific Gyre Oscillation modulates seasonal timing and ecosystem functioning in the California Current upwelling system. *Geophysical Research Letters* **39**(1).

Chhak K, Di Lorenzo E. 2007. Decadal variations in the California Current upwelling cells. *Geophysical Research Letters* **34**(14).

Countway PD, Caron DA. 2006. Abundance and distribution of Ostreococcus sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Applied and Environmental Microbiology* **72**(4): 2496–2506.

Di Lorenzo E, Combes V, Keister JE, Strub PT, Thomas AC, Franks PJ, Ohman MD, Furtado JC, Bracco A, Bograd SJ. 2013. Synthesis of Pacific Ocean climate and ecosystem dynamics. *Oceanography* **26**(4): 68–81.

Di Lorenzo E, Schneider N, Cobb KM, Franks PJS, Chhak K, Miller AJ, McWilliams JC, Bograd SJ, Arango H, Curchitser E. 2008. North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophysical Research Letters* **35**(8).

Dong C, Idica EY, McWilliams JC. 2009. Circulation and multiple-scale variability in the Southern California Bight. *Progress in Oceanography* **82**(3): 168–190. Elsevier.

Dong C, McWilliams JC, Hall A, Hughes M. 2011. Numerical simulation of a synoptic event in the Southern California Bight. *Journal of Geophysical Research: Oceans* **116**(C5). Wiley Online Library.

Dong, C., L. Renault, Y. Zhang, J, Ma, and Y. Cao, 2017: Expansion of West Coast Oceanographic Modeling Capability. US Department of the Interior, Bureau of Ocean Energy Management, Pacific. OCS Study BOEM 2017-055. 83 pp.

Fewings, M.R., Washburn, L. and Ohlmann, J.C., 2015. Coastal water circulation patterns around the northern Channel Islands and Point Conception, California. *Progress in Oceanography*, *138*, pp.283-304.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**(5374): 237–240.

Fischer AD, Hayashi K, McGaraghan A, Kudela RM. 2020. Return of the "age of dinoflagellates" in Monterey Bay: Drivers of dinoflagellate dominance examined using automated imaging flow cytometry and long-term time series analysis. *Limnology and Oceanography*. Wiley Online Library.

Goodman D, Eppley R, Reid F. 1984. Summer phytoplankton assemblages and their environmental correlates in the Southern California Bight. *Journal of Marine Research* **42**(4): 1019–1049. Sears Foundation for Marine Research.

Goodman J, Brzezinski MA, Halewood ER, Carlson CA. 2012. Sources of phytoplankton to the inner continental shelf in the Santa Barbara Channel inferred from cross-shelf gradients in biological, physical and chemical parameters. *Continental Shelf Research* **48**: 27–39.

Gregorio DE, Pieper RE. 2000. Investigations of red tides along the southern California coast. *Bulletin of the Southern California Academy of Sciences* **99**(3): 147–147.

Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum JR. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**(7600): 465.

Guillocheau, N, 2003. Beta-Correction Experiment Report. ICESS Internal document, UCSB, April 2003.

Hamed, KH, & Rao, AR. 1998. A modified Mann-Kendall trend test for autocorrelated data. *Journal of hydrology*, **204**(1-4), 182-196.

Harms S, Winant CD. 1998. Characteristic patterns of the circulation in the Santa Barbara Channel. *Journal of Geophysical Research: Oceans* **103**(C2): 3041–3065.

Henderikx Freitas F, Siegel DA, Maritorena S, Fields E. 2017. Satellite assessment of particulate matter and phytoplankton variations in the Santa Barbara Channel and its surrounding waters: R ole of surface waves. *Journal of Geophysical Research: Oceans* **122**(1): 355–371.

Higgins HW, Wright SW, Schluter L. 2011. Quantitative interpretation of chemotaxonomic pigment data. In: *Phytoplankton Pigments: Characterization, Chemotaxonomy, and Applications in Oceanography*. Cambridge University Press.

Hirata T, Hardman-Mountford N, Brewin R, Aiken J, Barlow R, Suzuki K, Isada T, Howell E, Hashioka T, Noguchi-Aita M. 2011. Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences* **8**(2): 311–327. Copernicus Publications.

Hooker SB, Thomas CS, Van Heukelem L, Russ ME, Ras J, Claustre H, Clementson L, Canuti E, Berthon J-F, Perl J. 2010. The fourth SeaWiFS HPLC analysis round-Robin experiment (SeaHARRE-4). NASA Technical Memorandum.

Jacox MG, Hazen EL, Zaba KD, Rudnick DL, Edwards CA, Moore AM, Bograd SJ. 2016. Impacts of the 2015–2016 El Niño on the California Current System: Early

assessment and comparison to past events. *Geophysical Research Letters* **43**(13): 7072–7080.

Jacox MG, Moore AM, Edwards CA, Fiechter J. 2014. Spatially resolved upwelling in the California Current System and its connections to climate variability. *Geophysical Research Letters* **41**(9): 3189–3196.

Jeffrey SW, Wright SW, Zapata M. 2011. Microalgal classes and their signature pigments. In: *Phytoplankton Pigments: Characterization, Chemotaxonomy, and Applications in Oceanography*. Cambridge University Press.

Johnson KS, Petty RL, Thomsen J. 1985. Flow-injection analysis for seawater micronutrients. *Advances in Chemistry* **209**. ACS Publications.

Kahru M, Kudela RM, Manzano-Sarabia M, Mitchell BG. 2012. Trends in the surface chlorophyll of the California Current: Merging data from multiple ocean color satellites. *Deep Sea Research Part II: Topical Studies in Oceanography* **77**: 89–98. Elsevier.

Kramer SJ, Siegel, DA. 2019. How can phytoplankton pigments be best used to characterize surface ocean phytoplankton groups for ocean color remote sensing algorithms?. *Journal of Geophysical Research: Oceans*, **124**(11), 7557-7574. Wiley Online Library.

Krause JW, Brzezinski MA, Siegel DA, Thunell RC. 2013. Biogenic silica standing stock and export in the Santa Barbara Channel ecosystem. *Journal of Geophysical Research: Oceans* **118**(2): 736–749. Wiley Online Library.

Latasa M, Bidigare RR. 1998. A comparison of phytoplankton populations of the Arabian Sea during the Spring Intermonsoon and Southwest Monsoon of 1995 as

described by HPLC-analyzed pigments. *Deep Sea Research Part II: Topical Studies in Oceanography* **45**(10–11): 2133–2170.

Lin Y, Cassar N, Marchetti A, Moreno C, Ducklow H, Li Z. 2017. Specific eukaryotic plankton are good predictors of net community production in the Western Antarctic Peninsula. *Scientific reports* **7**(1): 1–11. Nature Publishing Group.

Lombard F, Boss E, Waite AM, Vogt M, Uitz J, Stemmann L, Sosik HM, Schulz J, Romagnan J-B, Picheral M. 2019. Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science* **6**: 196. Frontiers.

Lucas AJ, Dupont CL, Tai V, Largier JL, Palenik B, Franks PJ. 2011. The green ribbon: Multiscale physical control of phytoplankton productivity and community structure over a narrow continental shelf. *Limnology and Oceanography* **56**(2): 611–626.

Mackey M, Mackey D, Higgins H, Wright S. 1996. CHEMTAX-a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Marine Ecology Progress Series* **144**: 265–283.

Mantua NJ, Hare S. 2002. Pacific-Decadal Oscillation (PDO). *Encyclopedia of global environmental change* **1**: 592–594.

Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the american Meteorological Society* **78**(6): 1069–1080.

Margalef R. 1978. Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanologica acta* **1**(4): 493–509. Gauthier-Villars.

Matson PG, Washburn L, Fields EA, Gotschalk C, Ladd TM, Siegel DA, Welch ZS, Iglesias-Rodriguez MD. 2019. Formation, development, and propagation of a rare coastal coccolithophore bloom. *Journal of Geophysical Research: Oceans* **124**(5): 3298–3316.

McPhee-Shaw EE, Siegel DA, Washburn L, Brzezinski MA, Jones JL, Leydecker A, Melack J. 2007. Mechanisms for nutrient delivery to the inner shelf: Observations from the Santa Barbara Channel. *Limnology and Oceanography* **52**(5): 1748–1766.

Melton C, Washburn L, Gotschalk C. 2009. Wind relaxations and poleward flow events in a coastal upwelling system on the central California coast. *Journal of Geophysical Research: Oceans* **114**(C11).

Mitarai S, Siegel DA, Watson J, Dong C, McWilliams J. 2009. Quantifying connectivity in the coastal ocean with application to the Southern California Bight. *Journal of Geophysical Research: Oceans* **114**(C10). Wiley Online Library.

Mitchell BG. 1990. Algorithms for determining the absorption coefficient for aquatic particulates using the quantitative filter technique. Ocean optics X **1302**: 137–148. International Society for Optics and Photonics.

Needham DM, Fuhrman JA. 2016. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nature Microbiology* **1**(4): 16005.

Nygaard K, Tobiesen A. 1993. Bacterivory in algae: a survival strategy during nutrient limitation. *Limnology and Oceanography* **38**(2): 273–279. Wiley Online Library.

Ohlmann JC, Mitarai S. 2010. Lagrangian assessment of simulated surface current dispersion in the coastal ocean. *Geophysical Research Letters* **37**(17). Wiley Online Library.

Otero MP, Siegel DA. 2004. Spatial and temporal characteristics of sediment plumes and phytoplankton blooms in the Santa Barbara Channel. *Deep Sea Research Part II: Topical Studies in Oceanography* **51**(10–11): 1129–1149.

Palenik B. 2000. Picophytoplankton seasonal cycle at the SIO pier, La Jolla, California. *Journal of Phycology* **36**: 53–53. Wiley Online Library.

Reid F, Stewart E, Eppley R, Goodman D. 1978. Spatial distribution of phytoplankton species in chlorophyll maximum layers off southern California. *Limnology and Oceanography* **23**(2): 219–226. Wiley Online Library.

Roesler C, Stramski D, D'Sa E, Röttgers R, Reynolds RA. 2018. Spectrophotometric measurements of particulate absorption using filter pads. *Washington, DC: IOCCG.*

Santoro AE, Nidzieko NJ, van Dijken GL, Arrigo KR, Boehma AB. 2010. Contrasting spring and summer phytoplankton dynamics in the nearshore Southern California Bight. *Limnology and Oceanography* **55**(1): 264–278.

Schulien, JA, Peacock, MB, Hayashi, K, Raimondi, P, Kudela, RM. 2017. Phytoplankton and microbial abundance and bloom dynamics in the upwelling shadow of Monterey Bay, California, from 2006 to 2013. *Marine Ecology Progress Series*, **572**: 43-56.

Shipe R, Passow U, Brzezinski M, Graham W, Pak D, Siegel D, Alldredge A. 2002. Effects of the 1997–98 El Nino on seasonal variations in suspended and sinking

particles in the Santa Barbara basin. *Progress in Oceanography* **54**(1–4): 105–127. Elsevier.

Siegel DA, Behrenfeld MJ, Maritorena S, McClain CR, Antoine D, Bailey SW, Bontempi PS, Boss ES, Dierssen HM, Doney SC. 2013. Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. *Remote Sensing of Environment* **135**: 77–91.

Simons RD, Nishimoto MM, Washburn L, Brown KS, Siegel DA. 2015. Linking kinematic characteristics and high concentrations of small pelagic fish in a coastal mesoscale eddy. *Deep Sea Research Part I: Oceanographic Research Papers* **100**: 34–47.

Simons RD, Siegel DA, Brown KS. 2013. Model sensitivity and robustness in the estimation of larval transport: a study of particle tracking parameters. *Journal of Marine Systems* **119**: 19–29. Elsevier.

Smayda TJ, Trainer VL. 2010. Dinoflagellate blooms in upwelling systems: Seeding, variability, and contrasts with diatom bloom behaviour. *Progress in Oceanography* **85**(1–2): 92–107.

Taylor AG, Landry MR. 2018. Phytoplankton biomass and size structure across trophic gradients in the southern California Current and adjacent ocean ecosystems. *Marine Ecology Progress Series* **592**: 1–17.

Taylor AG, Landry MR, Selph KE, Wokuluk JJ. 2015. Temporal and spatial patterns of microbial community biomass and composition in the Southern California Current Ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanography* **112**: 117–128.

Thomson RE, Emery WJ. 2014. *Data Analysis Methods in Physical Oceanography*. Newnes.

Trees CC, Kennicutt II MC, Brooks JM. 1985. Errors associated with the standard fluorimetric determination of chlorophylls and phaeopigments. *Marine Chemistry* **17**(1): 1–12. Elsevier.

Uitz J, Claustre H, Morel A, Hooker SB. 2006. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research: Oceans* **111**(C8).

U.S. Integrated Ocean Observing System. 2013. Manual for Real-Time Quality Control of In-situ Temperature and Salinity Data: A Guide to Quality Control and Quality Assurance of In-situ Temperature and Salinity Observations.

Van Boekel J, Stefels W. 1993. Production of DMS from dissolved DMSP in axenic cultures of the marine phytoplankton species Phaeocystis sp. *Marine Ecology Progress Series* **97**: 11–18.

Van Heukelem L, Thomas CS. 2001. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *Journal of Chromatography A* **910**(1): 31–49.

de Vargas C, Aubry M-P, Probert I, Young J. 2007. Origin and evolution of coccolithophores: from coastal hunters to oceanic farmers. In: *Evolution of Primary Producers in the Sea*. Elsevier. p. 251–285.

Venrick EL. 1998. The phytoplankton of the Santa Barbara Basin: patterns of chlorophyll and species structure and their relationships with those of surrounding stations.

California Cooperative Oceanic Fisheries Investigations Report: 124–132. California Cooperative Oceanic Fisheries Investigations.

Venrick EL. 2002. Floral patterns in the California Current System off southern California: 1990-1996. *Journal of Marine Research* **60**(1): 171–189.

Venrick EL. 2012. Phytoplankton in the California Current system off southern California: Changes in a changing environment. *Progress in Oceanography* **104**: 46–58. Elsevier.

Vidussi F, Claustre H, Manca BB, Luchetta A, Marty J. 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *Journal of Geophysical Research: Oceans* **106**(C9): 19939–19956. Wiley Online Library.

Warrick JA, Mertes LA, Washburn L, Siegel DA. 2004. A conceptual model for river water and sediment dispersal in the Santa Barbara Channel, California. *Continental Shelf Research* **24**(17): 2029–2043. Elsevier.

Wear EK, Carlson CA, James AK, Brzezinski MA, Windecker LA, Nelson CE. 2015. Synchronous shifts in dissolved organic carbon bioavailability and bacterial community responses over the course of an upwelling-driven phytoplankton bloom. *Limnology and Oceanography* **60**(2): 657–677.

Winant CD, Dever EP, Hendershott MC. 2003. Characteristic patterns of shelf circulation at the boundary between central and southern California. *Journal of Geophysical Research: Oceans* **108**(C2). Wiley Online Library.

Wolter K, Timlin MS. 1993. Monitoring ENSO in COADS with a Seasonally Adjusted Principal. 1993. Proc. of the 17th Climate Diagnostics Workshop, Norman, OK,

NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma, 52.

Worden AZ, Nolan JK, Palenik B. 2004. Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. *Limnology and oceanography* **49**(1): 168–179.

Zapata M, Fraga S, Rodríguez F, Garrido JL. 2012. Pigment-based chloroplast types in dinoflagellates. *Marine Ecology Progress Series* **465**: 33–52.

Figures and Tables

Table 1. Pigment abbreviations and biomarker assumptions used in the present study. The five representative biomarker pigments and their taxonomic representation were inferred from the results of the cluster analysis presented in Figure 2 and the literature (Vidussi et al., 2001; Uitz et al., 2006; Jeffrey et al., 2011). The color-coding of each biomarker pigment corresponds to that used in subsequent figures.

Pigment	Abbreviation	Assumed Taxonomic Significance
Total chlorophyll *a*	TChla[1]	All phytoplankton
Total chlorophyll *b*	TChlb	Chlorophytes
Alpha-beta-carotene	ABCar	-
19'-butanoyloxyfucoxanthin	But	-
19'-hexanoyloxyfucoxanthin	Hex	Prymnesiophytes
Alloxanthin	Allo	-
Diadinoxanthin	Diadino	-
Diatoxanthin	Diato	-
Fucoxanthin	Fuco	Diatoms
Peridinin	Perid	Dinoflagellates
Zeaxanthin	Zea	Picophytoplankton
Divinyl chlorophyll *a*	DVChla	-
Chlorophyll c1 + c2	Chlc1c2	-
Chlorophyll c3	Chlc3	-
Lutein	Lut	-
Neoxanthin	Neo	-
Violaxanthin	Viola	-
Prasinoxanthin	Pras	-

[1]TChla is used to indicate total chlorophyll *a* concentrations determined by HPLC. Because the bio-optical models are trained on HPLC data, TChla also includes bio-optically modeled concentrations of total chlorophyll *a*. CHL is used to denote chlorophyll *a* concentrations determined by fluorometric methods (see Section 2.2).

Table 2. Selected mean (standard deviation) goodness of fit statistics from the 500-fold model cross-validation procedure. See Supporting Table S1 for a more complete listing of goodness of fit statistics.

Pigment	R^2	Median % error
TChla	0.87 (0.07)	17.2 (2.18)
TChlb	0.86 (0.04)	21.7 (2.51)
Hex	0.72 (0.06)	29.8 (3.69)
Fuco	0.87 (0.07)	35.0 (4.93)
Perid	0.88 (0.05)	98.8 (3.68)
Zea	0.54 (0.09)	38.3 (3.72)

Table 3. Summary statistics of the merged biomarker pigment data set. All stations are considered independently. Max, maximum; CV, coefficient of variation (standard deviation divided by the mean); r, Pearson's correlation coefficient. Insignificant correlations at 95% confidence are not shown.

Pigment	Mean $\mu g\ L^{-1}$	Median $\mu g\ L^{-1}$	Max $\mu g\ L^{-1}$	CV (%)	r, TChlb	r, Hex	r, Fuco	r, Perid	r, Zea
TChla	2.57	1.53	35.0	122	0.27	0.09	0.95	0.28	
TChlb	0.11	0.08	0.65	84.9		0.65	0.12	0.19	0.51
Hex	0.15	0.12	1.13	80.4				0.10	0.36
Fuco	0.95	0.41	19.2	172					-0.07
Perid	0.16	0.05	7.13	264					0.10
Zea	0.05	0.04	0.28	71.4					

Table 4. Pearson's correlation coefficients amongst normalized pigment anomaly time series (see Figure 9). Insignificant correlations ($p > 0.05$) are not shown.

Pigment	TChlb	Hex	Fuco	Perid	Zea
TChla	0.29		0.94	0.43	
TChlb		0.71		0.17	0.65
Hex			-0.14	0.17	0.45
Fuco				0.19	
Perid					0.17

Table 5. Correlation coefficients of pigment concentration and pigment-oceanographic EOF mode amplitude anomalies with climate forcings. Pigment concentration anomalies are normalized to climatological mean values as in Figures 9 and 12. Insignificant correlation coefficients ($p < 0.05$) are not shown.

	SOI	MEI	NPGO	PDO
TChla	0.31	-0.27		-0.19
TChlb				
Hex				
Fuco	0.30	-0.26		-0.20
Perid			-0.22	
Zea				
Mode 1	0.20	-0.23	0.24	-0.25
Mode 2		0.19	-0.16	0.19
Mode 3	0.19		-0.26	
Mode 4			0.16	

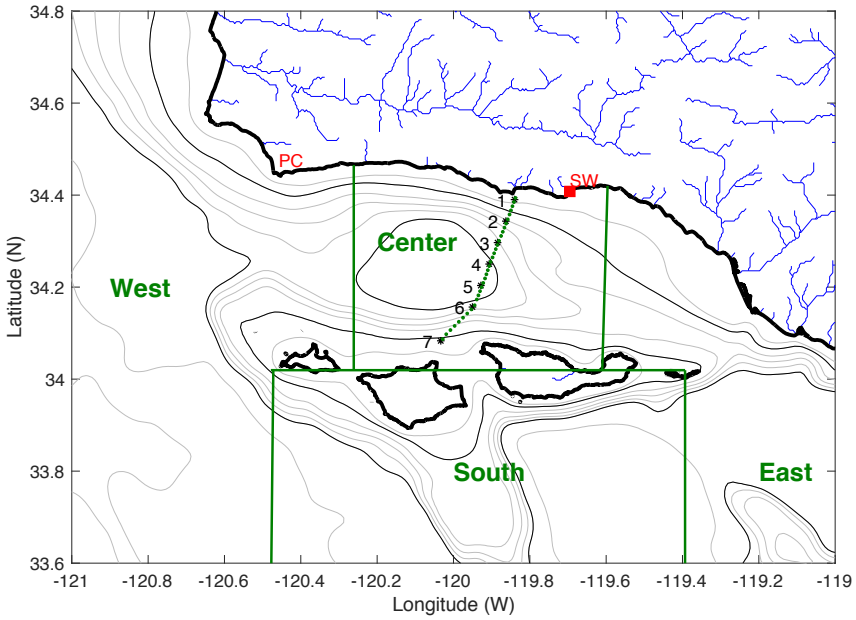

Figure 1. Map of the Santa Barbara Channel, CA. Plumes and Blooms stations are marked with black stars and labeled with station numbers. Bold black lines indicate coastline. Gray bathymetry contours are shown at 50, 200, 300, 400, 1000, 2000, 3000, and 4000 m water depth. Black bathymetry contours are the 100 and 500 m isobaths. Particle release points used in the ROMS particle tracking model are shown with green circles, and the four "origin boxes" (West, Center, East, South) used to define Santa Barbara Channel source waters are outlined with bold green lines (see Sections 2.8.2 and 4.3). The red square indicates Stearns Wharf (SW) where weekly microscopic counts of several phytoplankton species are performed (see Section 2.8.3). Point Conception (PC) is noted in red.

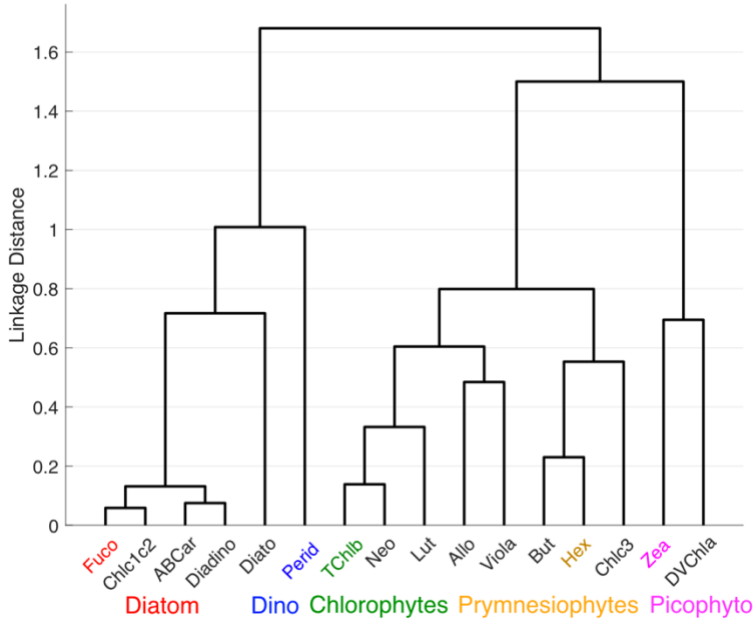

Figure 2. Hierarchical cluster analysis of HPLC phytoplankton pigment concentrations using the correlation distance and Ward's linkage method. Representative biomarker pigments and phytoplankton groups are color-coded here and in all subsequent analysis to aid interpretation.

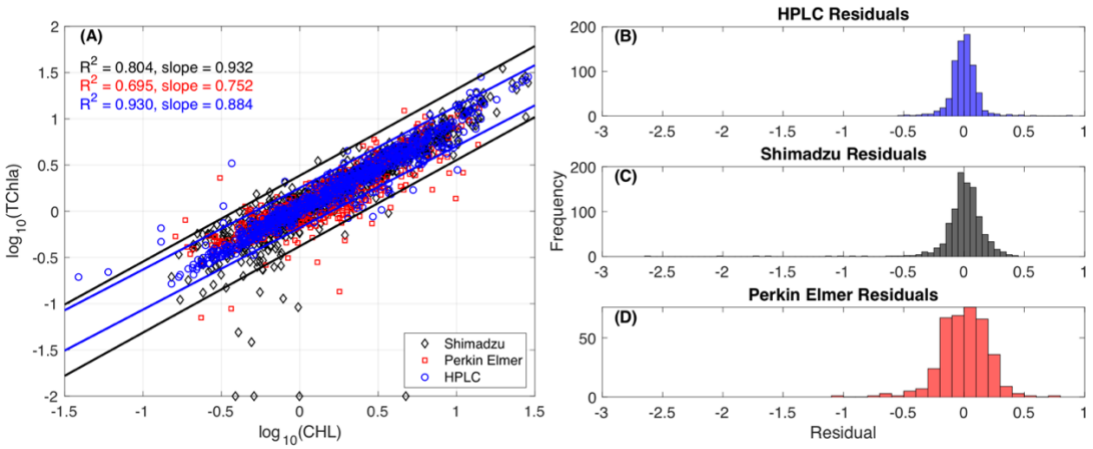

Figure 3. Regression (A) and residual (B, C, D) analysis of HPLC TChla concentrations (blue), TChla concentrations modeled from the Shimadzu 2401-PC $a_{ph}(\lambda)$ (black), and modeled TChla concentrations derived from the Perkin-Elmer Lambda 2 $a_{ph}(\lambda)$ (red) against fluorometric chlorophyll *a* concentrations (CHL). Blue and black lines in the scatterplot are 95% prediction intervals for the HPLC and Shimadzu TChla regressions, respectively. All residuals are log-transformed. 10^{-2} µg L^{-1} was added to all values to prevent undefined values in the log-transformed data.

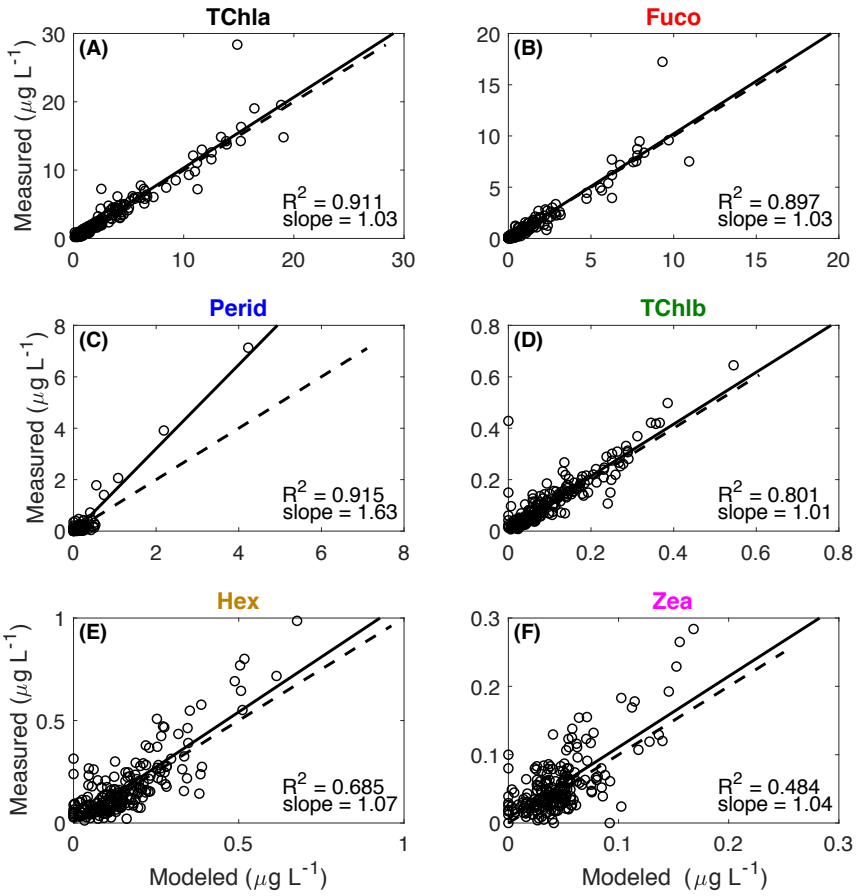

Figure 4. Validation of bio-optical models for (A) TChla, (B) Fuco, (C) Perid, (D) TChlb, (E) Hex, and (F) Zea extrapolated to HPLC observations not included in the model cross-validation exercise. Panel titles are color-coded as in Figure 2.

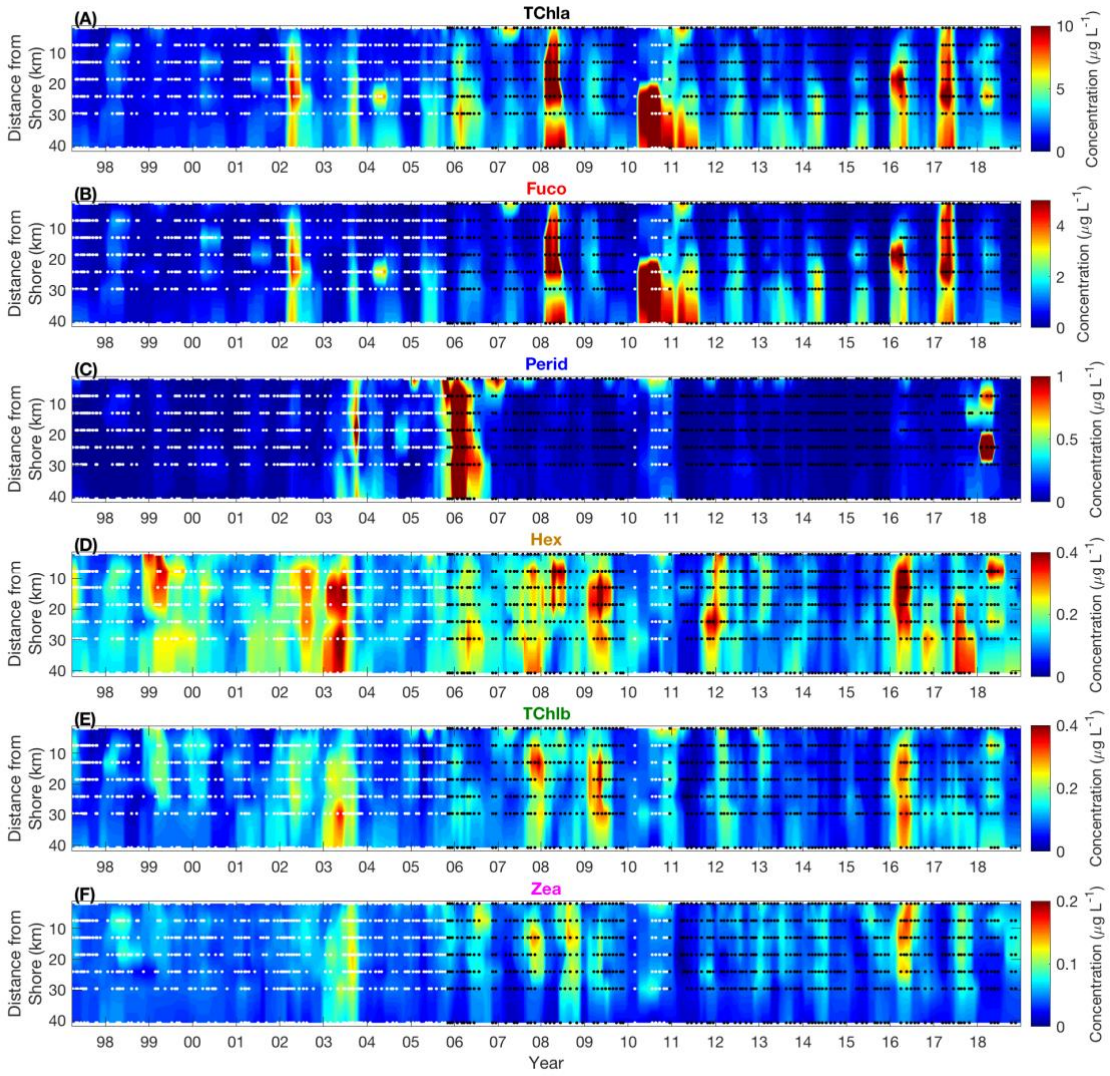

Figure 5. Spatiotemporal dynamics of (A) TChla, (B) Fuco, (C) Perid, (D) Hex, (E) TChlb, and (F) Zea. The top of each panel corresponds to PnB station 1 on the mainland shelf. White dots show modeled pigment concentrations, while black dots show pigment concentrations measured by HPLC. Ordinary kriging with an exponential-Bessel fitting model (GLOBEC Kriging Software Package v3.0) was used to smooth the data for this figure. Interpolation length scales are 30 days and 5 km in the cross-shelf direction. Panel titles are color-coded as in Figure 2.

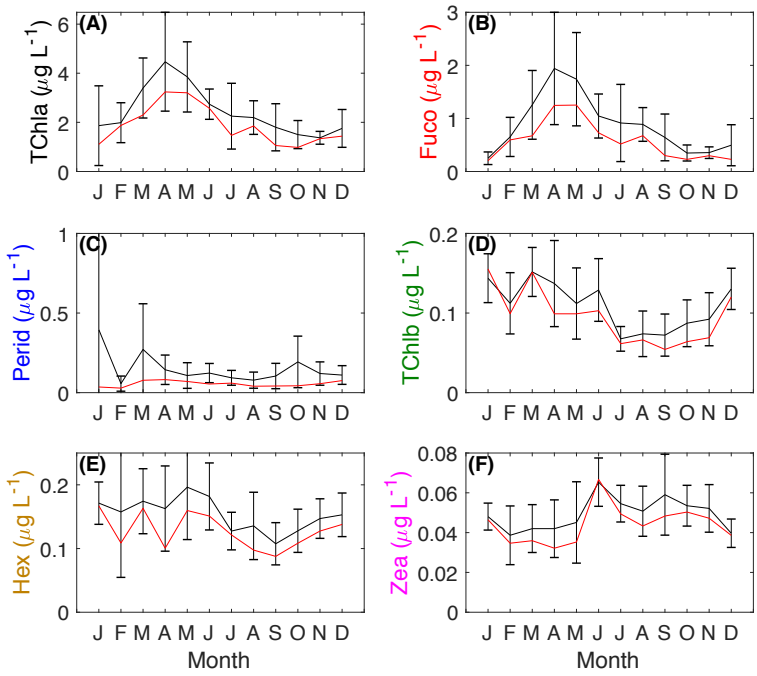

Figure 6. Mean +/- 95% confidence intervals (black) and median (red) annual cycles of (A) TChla, (B) Fuco, (C) Perid, (D) TChlb, (E) Hex, and (F) Zea. Annual cycles were computed based on each pigments monthly mean time series determined by averaging each pigments' concentrations by sampling event and when more than one sampling event occurred in the same month, by month. Y-axis labels are color-coded as in Figure 2.

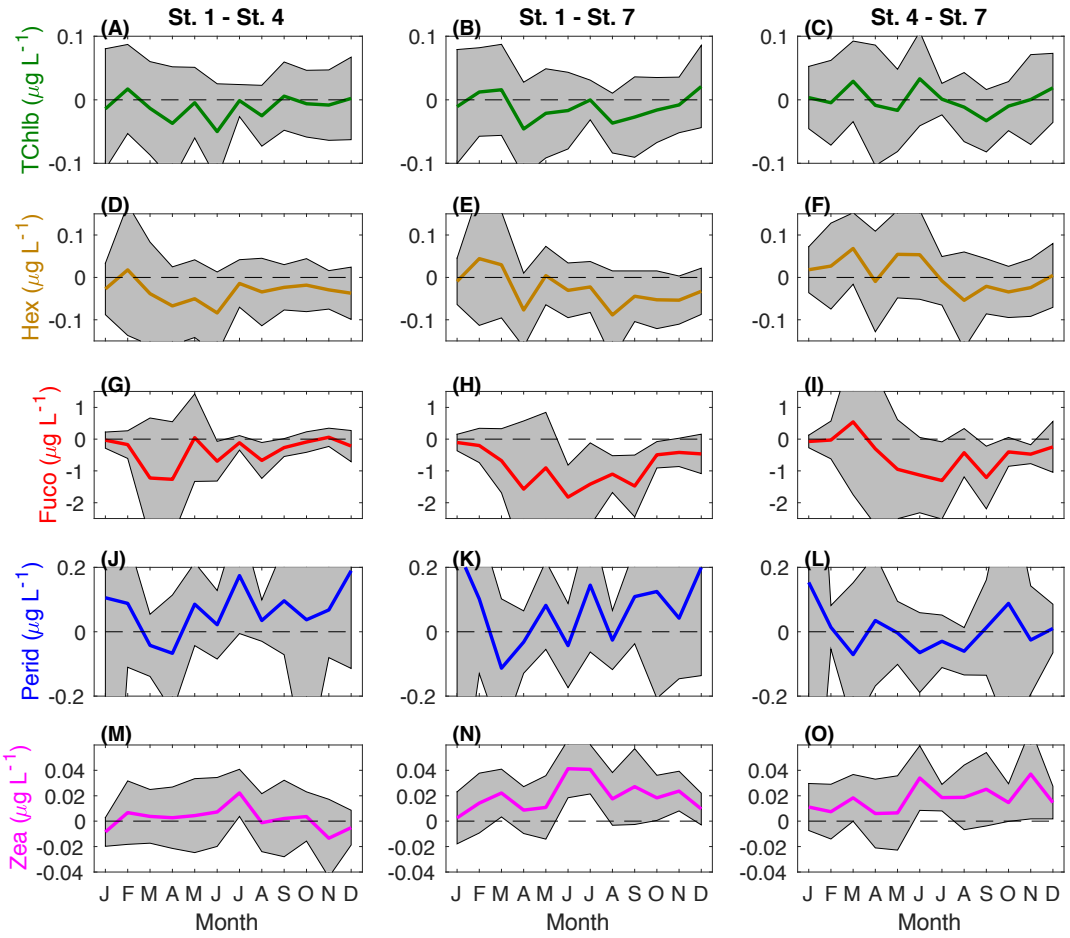

Figure 7. Spatial differences in mean annual cycles across PnB stations (A, D, G, J, M) 1 and 4, (B, E, H, K, N) 1 and 7, and (C, F, I, L, O) 4 and 7, for the five major biomarker pigment concentrations. The shaded region around each line corresponds to the 95% confidence interval computed for the difference of monthly mean pigment concentrations. Significant differences in monthly mean pigment concentrations at 95% confidence thus occur where the shaded region does not overlap the dashed zero line. Y-axis labels and lines are color-coded as in Figure 2.

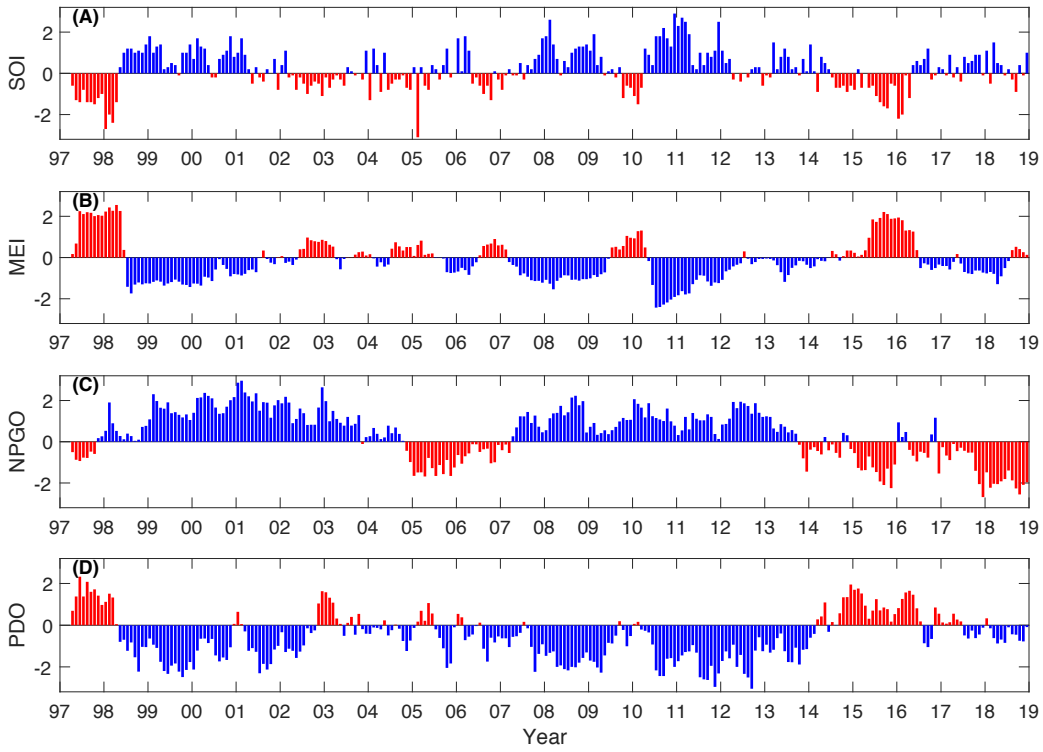

Figure 8. Indices of the dominant modes of North Pacific climate variability, including the (A, B) El Niño Southern Oscillation represented by the (A) Southern Oscillation Index (SOI) and the (B) Multivariate ENSO Index (MEI), (C) the North Pacific Gyre Oscillation (NPGO), and (D) the Pacific Decadal Oscillation (PDO). Blue and red bars indicate "cold" and "warm" phases, respectively.

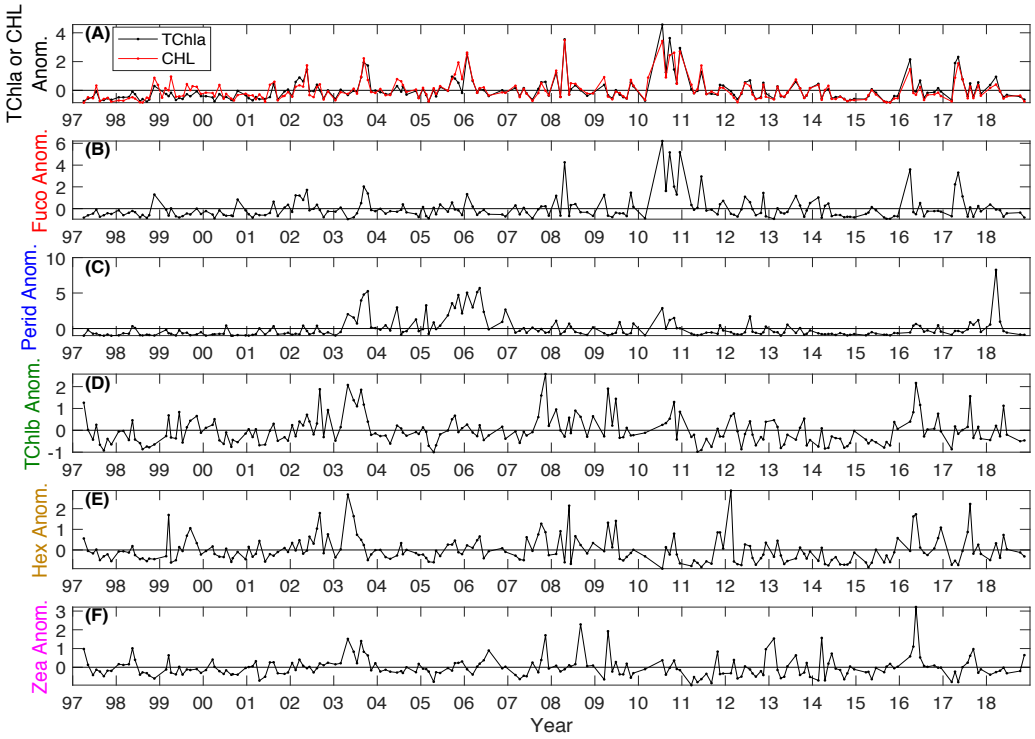

Figure 9. Normalized seasonal anomaly time series for (A) TChla (black) and fluorometric CHL (red), (B) Fuco, (C) Perid, (D) TChlb, (E) Hex, (F) Zea. Anomalies are computed by subtracting the climatological mean pigment concentrations shown in Figure 6 from each pigment concentration's monthly mean time series, and then normalizing to the climatological mean pigment concentration. Anomalies are thus unitless and represent a fold-change from the annual cycles shown in Figure 6. Y-axis labels are color-coded as in Figure 2.

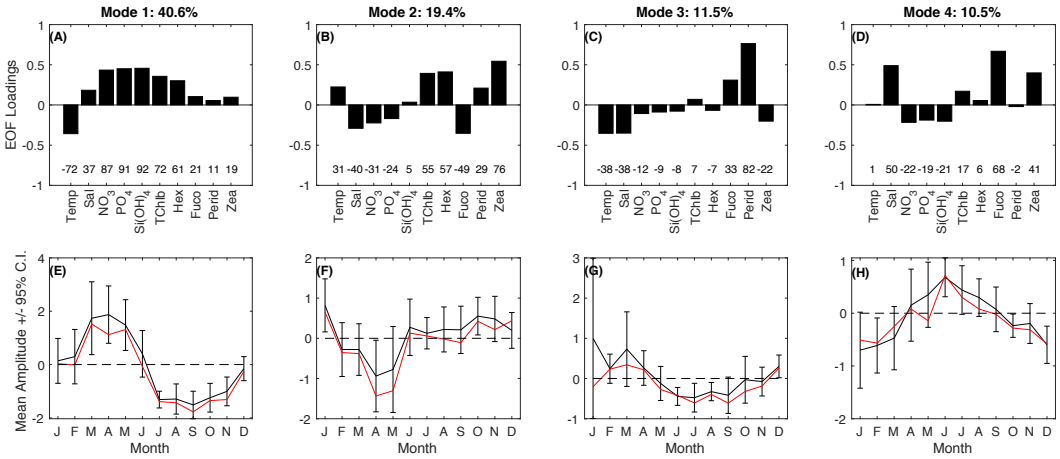

Figure 10. (A, B, C, D) Loadings and (E, F, G, H) mean +/- 95% confidence intervals (black) and median (red) annual cycles of the amplitude functions of the first four EOF modes of the pigment-oceanographic data set. The variance explained by each mode is indicated in the title of each panel. Numbers above each x-tick are the correlation coefficient between a particular EOF mode amplitude function and variable multiplied by 100. All variables were averaged by sampling event (spatially) and, where multiple cruises occurred in the same month, by month, to create a monthly time series of each variable. Monthly time series were then z-scored prior to computing EOFs.

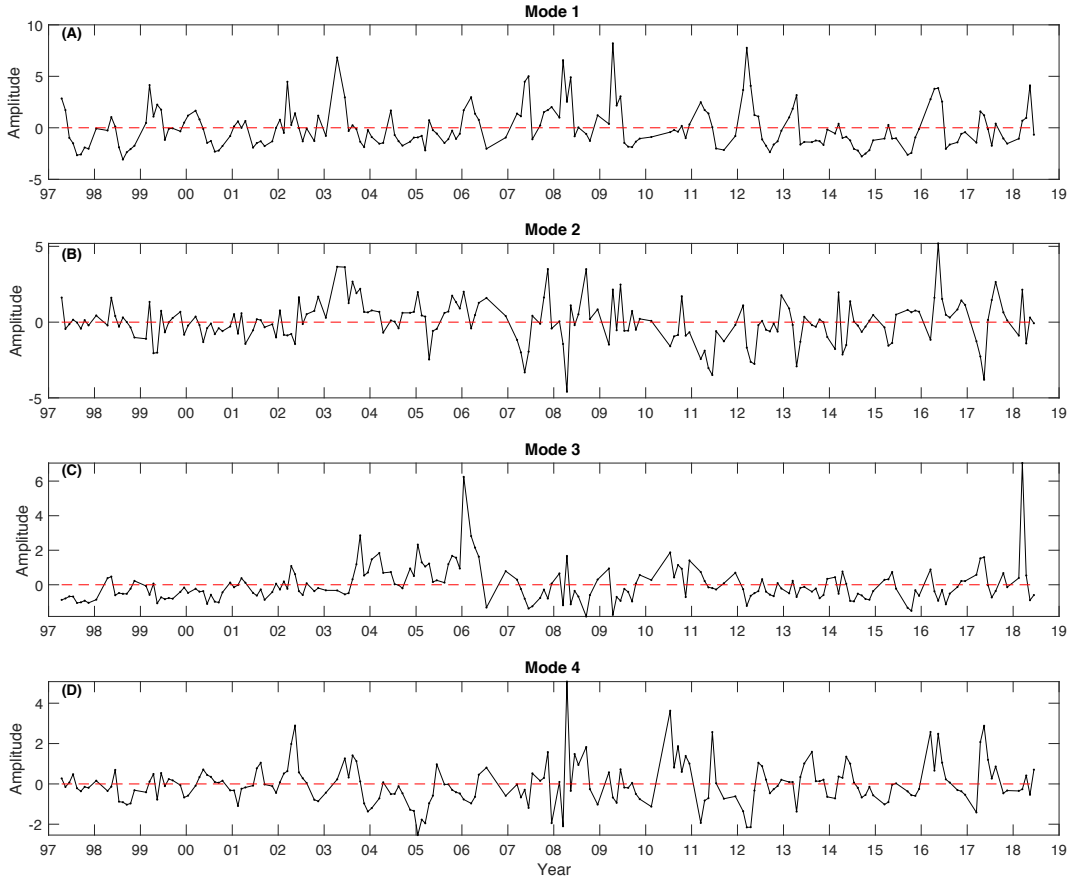

Figure 11. Amplitude time series of the first four pigment-oceanographic EOF modes.

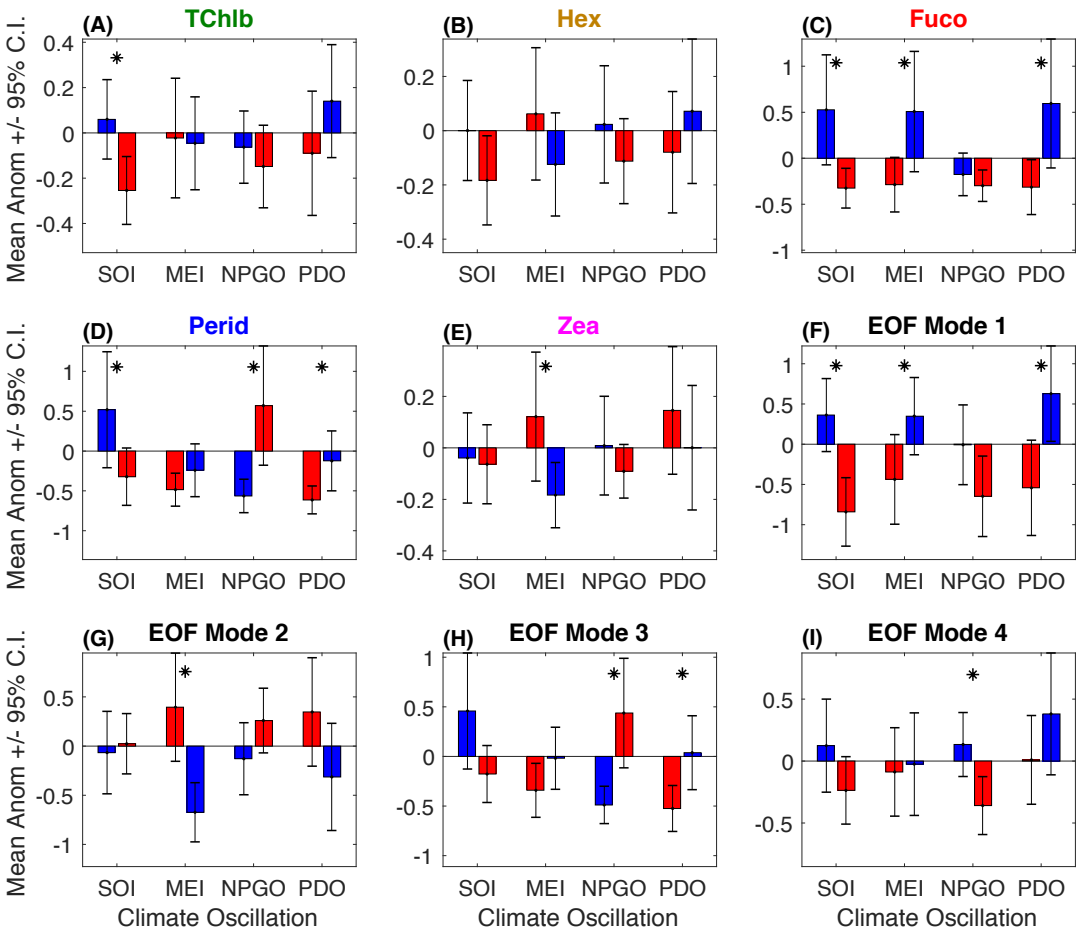

Figure 12. Conditional mean +/- 95% confidence intervals of biomarker pigment concentration and EOF amplitude anomalies according to the 15% largest positive and negative values of the Southern Oscillation Index (SOI), Multivariate ENSO Index (MEI), North Pacific Gyre Oscillation index (NPGO), and the Pacific Decadal Oscillation index (PDO). Red and blue bars indicate the "warm" and "cold" phases of each climate oscillation. Pigment concentration anomalies are normalized to climatological mean values, while EOF mode amplitude anomalies are not as all variables were standardized prior to the EOF analysis. Stars above each bar group indicate statistically significant ($p < 0.05$) differences across the warm and cold phase of the climate oscillation index according to a two-sample t-test. Panel titles are color-coded as in Figure 2.

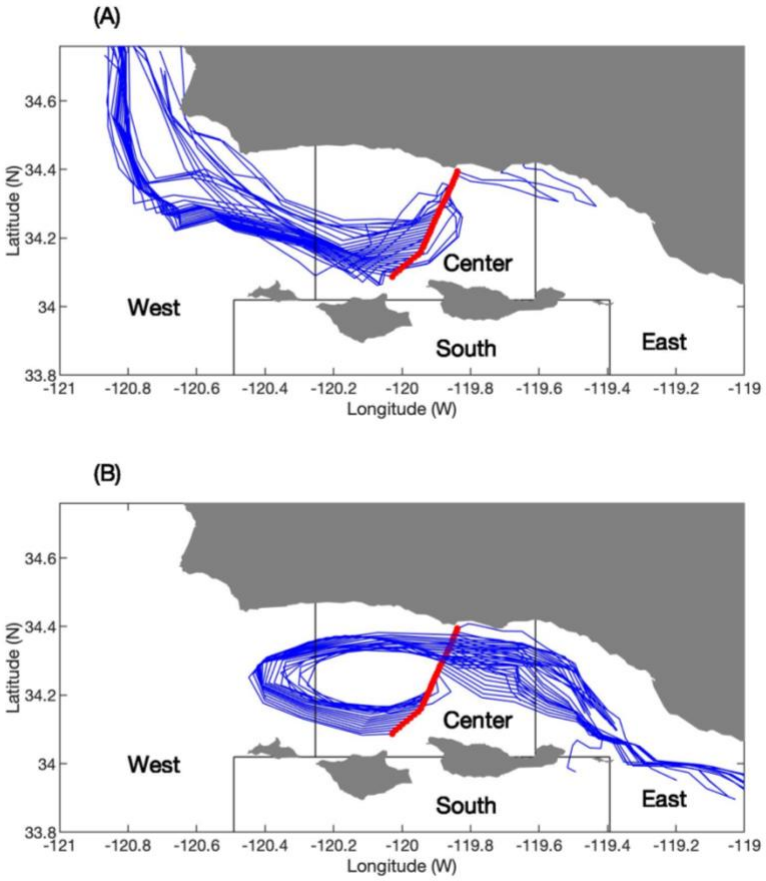

Figure 13. Example 10-day reverse-tracking particle trajectories simulated by the ROMS particle tracking model on (A) April 3, and (B) April 23, 2005. (A) shows typical particle trajectories during an upwelling event, with most particles originating from the West origin box, while (B) shows an upwelling-wind relaxation event driving higher advection of particles from the East origin box.

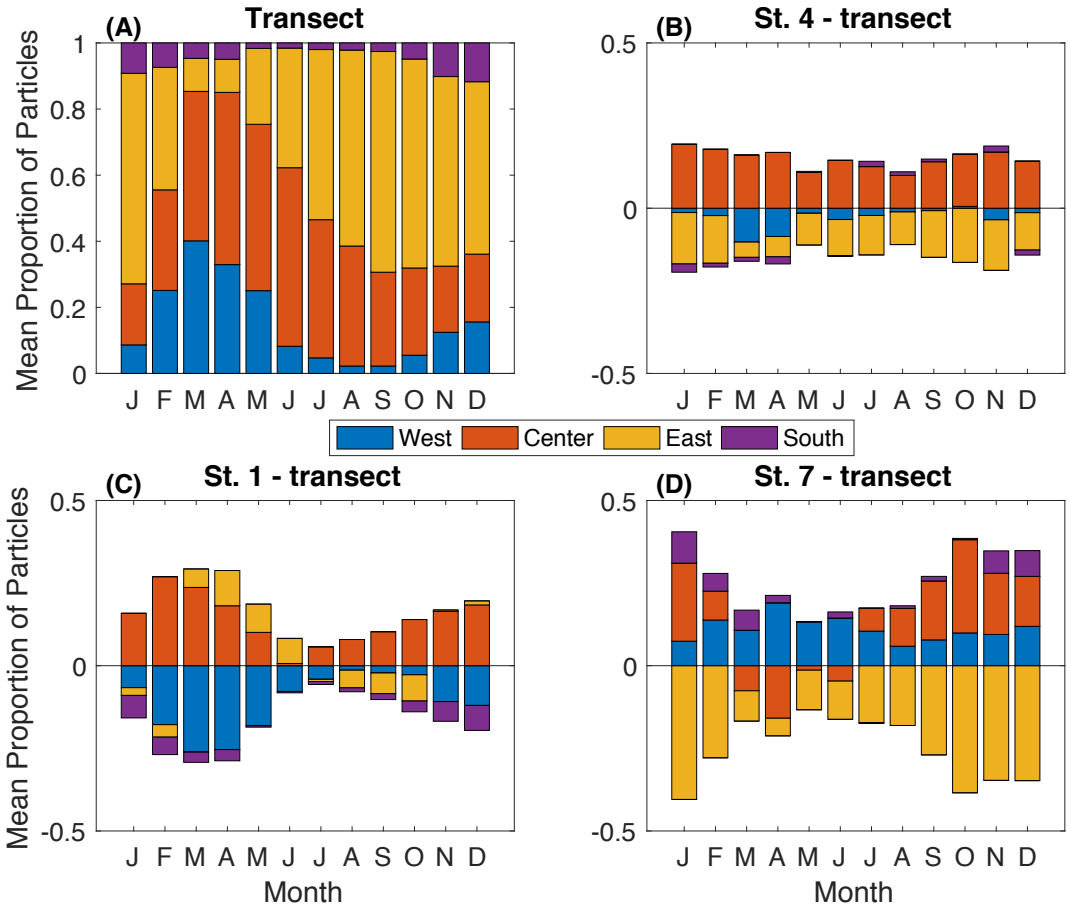

Figure 14. Mean annual cycles in the proportion of particles originating from each of the four origin boxes (see Figures 1 and 13) for (A) all release points on the PnB transect, and for the four to five release points closest to PnB stations (B) 4, (C) 1, and (D) 7 minus the transect mean annual cycle. The station-specific mean annual cycles are presented as differences relative to the entire transects mean annual cycle. The proportion of particles originating from the west (east) origin box serves as a proxy for the magnitude of advection of CCS (SCBight) source waters.

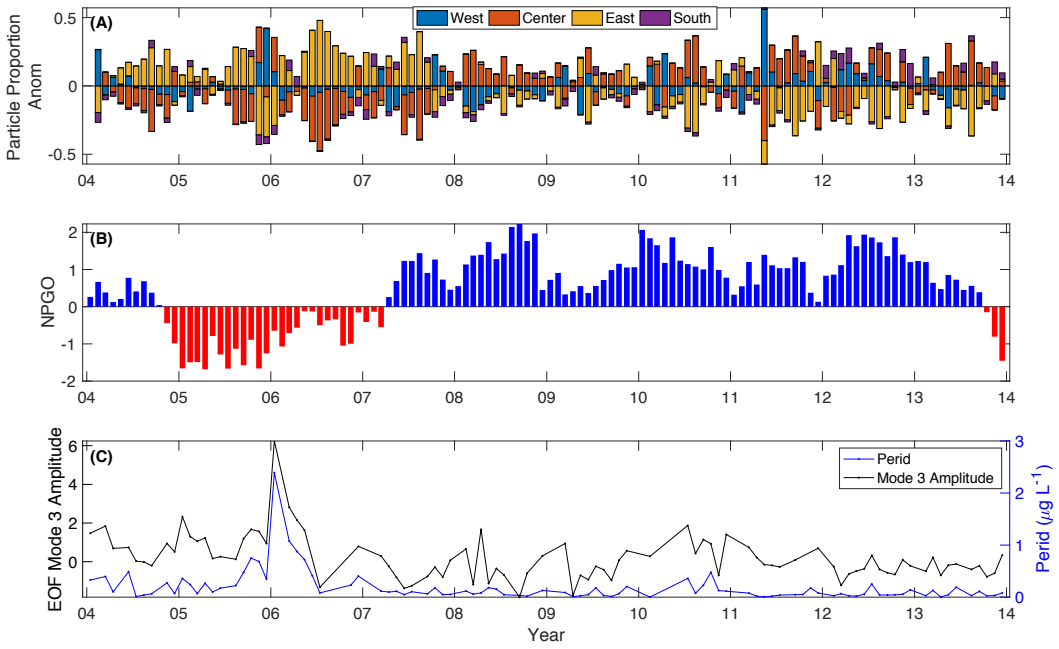

Figure 15. Time series of (A) anomalies in the proportion of particles advected from each of the four origin boxes determined by the ROMS particle tracking model (see Figures 1, 13, and 14) for the entire PnB transect, (B) the North Pacific Gyre Oscillation index, and (C) monthly mean Perid concentrations and EOF Mode 3 amplitudes.

IV. Integrating phytoplankton pigment and DNA meta-barcoding observations to determine phytoplankton community composition in the coastal ocean

Dylan Catlett, David A. Siegel, Paul G. Matson, Emma K. Wear, Craig A. Carlson, Thomas S. Lankiewicz, M. Debora Iglesias-Rodriguez

Abstract

Quantifying phytoplankton community composition (PCC) is critical to understanding and predicting marine ecosystem structure and function. DNA meta-barcoding and high performance liquid chromatography (HPLC) pigment analysis are two of the most widely applied methods for assessing PCC. While both of these methods are now applied on local to global scales, integrating these methods to assess PCC is rarely performed despite the need to coherently quantify PCC and the potential to gain novel insights into phytoplankton ecology. Here we integrate DNA meta-barcoding and HPLC pigment observations to assess phytoplankton composition in the Santa Barbara Channel, California, a highly productive coastal site. We find that both methods identify the same four dominant phytoplankton groups (diatoms, dinoflagellates, chlorophytes, and prymnesiophytes), but inter- and intra-lineage variability in biomarker pigment expression drives substantial disagreement between the two methods. Covariation network analysis circumvents uncertainty introduced by inter- and intra-lineage variability in pigment expression and reveals that diverse communities of phytoplankton and other protists covary with distinct suites of phytoplankton biomarker pigment concentrations. Our results highlight the strengths and weaknesses of each method in characterizing PCC, and reveal novel insights into phytoplankton physiology and ecology that could only be gained by integrating the two methods. Finally, we suggest a path to monitor marine ecosystem structure and function on unprecedented spatiotemporal scales based on the covariation of unique phytoplankton and protistan communities and food web interactions with remotely sensible phytoplankton pigment concentrations.

Introduction

Phytoplankton production fuels most marine food webs and the transport of carbon from the atmosphere to the ocean interior via the biological carbon pump (Ryther 1969; Guidi et al. 2016). The composition of the phytoplankton community determines the efficiency of the biological carbon pump and of the transfer of phytoplankton production to higher trophic levels, and thus dictates marine ecosystem structure and function (Ryther 1969; Trudnowska et al., 2021). Quantifying phytoplankton community composition (PCC) and/or stocks of particular phytoplankton groups (PGs) on regional to global scales thus represents a critical step in efforts to understand and predict marine ecosystem structure and function.

Quantifying PCC is immensely difficult due to the high genetic, morphological, taxonomic, and functional diversity of phytoplankton in the world's oceans (De Vargas et al. 2015). Various methods with unique strengths and weaknesses are available to quantify PCC and/or particular PGs including high performance liquid chromatography (HPLC) pigment analysis, high-throughput amplicon sequencing of phylogenetically informative DNA "barcode" genes (DNA meta-barcoding), various other "meta-omics" techniques (meta-genomics, transcriptomics, etc.), flow cytometry, and microscopic and/or image-based cell identification and enumeration (Lombard et al. 2019). Bio-optical and ocean color remote sensing approaches for estimating PCC have also been proposed but rely on other PCC methods (typically HPLC pigment analysis) for formulation and validation (Uitz et al., 2015; Chase et al., 2017; Catlett and Siegel 2018). Both HPLC pigment and DNA meta-barcoding methods entail relatively efficient sample analysis procedures, and are now widely applied to assess PCC on regional to global

scales (Lima-Mendez et al. 2015; Kramer and Siegel 2018). Recent work suggests that HPLC and DNA meta-barcoding analyses tend to provide estimates of PCC that are more comparable to one another than to other PCC methods, in part because they both provide relatively holistic characterizations of PCC spanning a broad range of size classes (Gong et al. 2020).

DNA meta-barcoding analysis results in a collection of operational taxonomic units or amplicon sequence variants (ASVs), both of which are typically assumed to represent "species" in microbial ecology applications (Callahan et al., 2016). Sequence counts for each ASV are quantified as the frequency of detection of the ASV sequence in each sample, although the total sequence counts recovered is arbitrarily constrained by sample preparation and/or analysis procedures (Gloor et al., 2017). Most DNA meta-barcoding data are thus compositional with ASV relative sequence abundances (RSAs) often estimated by normalizing each ASV's sequence counts to the total sequence counts recovered from each sample (e.g., De Vargas et al. 2015; Needham and Fuhrman, 2016). Standard Pearson and Spearman correlation analyses (and analyses reliant on these statistics) of compositional data can lead to spurious interpretations of ecological dynamics because variability in a single ASV's RSA can be driven either by changes in the abundance of other ASVs in the composition, or by a change in the abundance of the ASV in question (Aitchison, 1982; Gloor et al., 2017). While approaches that estimate ASV concentrations in situ are becoming popular (Lin et al., 2019), the compositionality constraint inherent in most DNA meta-barcoding data is a major obstacle in determining the links between oceanographic, geochemical, and PCC variability.

Wet lab and bioinformatic procedures for DNA meta-barcoding analysis vary widely across investigators, and in general analytical uncertainty in DNA meta-barcoding estimates of PCC is poorly constrained and difficult to evaluate (Wear et al., 2018; Catlett et al., 2020a). The choice of barcode gene and/or hypervariable region, PCR primers, and many other factors can alter DNA meta-barcoding estimates of PCC (Bradley et al., 2016; Wear et al., 2018). However, recent work suggests rigorously evaluated amplicon sequencing workflows provide reasonably accurate and precise estimates of PCC (Catlett et al., 2020a; Yeh et al., 2021). Further, current bioinformatic methods perform well in determining ASV sequences (Callahan et al. 2016) and in assigning taxonomy to ASV sequences (Murali et al., 2018). Despite their limitations, these data offer relatively high taxonomic resolution of PCC and are increasingly used to quantitatively assess the roles of phytoplankton communities in marine ecosystems and biogeochemical cycles (Guidi et al. 2016; Lin et al., 2017; Caputi et al. 2018). However, DNA meta-barcoding estimates of PCC have been conducted largely independently of other methods and evaluations of the relationships between DNA meta-barcoding and other methods' PCC estimates are urgently needed.

HPLC pigment analysis is also widely applied to estimate PCC in the world's oceans and is used almost exclusively to validate satellite ocean color PCC algorithms (Bracher et al., 2015; Chase et al. 2017; Kramer and Siegel 2018). HPLC analysis quantifies the concentrations of a suite of phytoplankton accessory pigments, some of which serve as biomarkers for particular PGs (Jeffrey et al. 2011). The primary weaknesses in assessing PCC with HPLC pigment analysis are (1) limited taxonomic resolution provided by accessory pigment data (to approximately the class level; e.g.,

Mackey et al. 1996; Kramer and Siegel 2018); (2) variability in pigment concentrations due to differences in physiological status and other factors in addition to changes in phytoplankton biomass (Goericke and Montoya, 1998; Schlüter et al., 2000); and (3) lack of specificity of commonly used biomarker pigments (Jeffrey et al. 2011). Widely used pigment chemotaxonomy methods include CHEMTAX (Mackey et al. 1996) and diagnostic pigment analysis (Vidussi et al. 2001; Uitz et al. 2006). However, no method is capable of accounting for the impacts of physiological variability or biomarker ambiguity on PCC estimates where only HPLC pigment data are available (Goericke and Montoya, 1998; Irigoien et al., 2004). Most chemotaxonomic methods impose compositionality on HPLC data by normalizing diagnostic pigment concentrations to the total chlorophyll a (TChla) concentration or the total concentration of all diagnostic pigments (Mackey et al. 1996; Vidussi et al. 2001; Uitz et al. 2006). Similar to compositional DNA meta-barcoding data, analysis of compositional pigment biomass proportions complicates interpretations of analyses that rely on Pearson or Spearman correlation (Aitchison, 1982; Catlett and Seigel, 2018). Despite these drawbacks, HPLC sample analysis methods are the most rigorously evaluated and standardized of any available PCC method (Van Heukelem and Thomas 2001), and the direct links between phytoplankton pigments and bio-optical properties provide a path to observe PCC quasi-continuously on global scales via satellite ocean color (Bracher et al., 2015; Chase et al., 2017; Catlett and Siegel, 2018).

Recent work suggests phytoplankton pigment “communities” can be identified from HPLC pigment data on regional to global scales based on the covariation of PGs and the concentrations of their accessory pigments with one another (Latasa and Bidigare,

1998; Catlett and Siegel, 2018; Kramer and Siegel, 2019). Specific biomarker pigments can then be selected as representatives of different phytoplankton communities, bio-optically modeled with relatively high accuracy, and subject to further analyses to determine oceanographic forcings of PGs on large spatiotemporal scales (Catlett et al., 2021a). The underlying principles of phytoplankton pigment community analysis are similar to those employed in recent studies that take a systems-level approach to characterize covariation amongst microbial ASVs using ecological networks, which can then be analyzed further to elucidate relationships between microbial communities and ecosystem structure and function (Lima-Mendez et al. 2015; Berdjeb et al. 2018). Given the numerous approaches for assessing PCC from HPLC pigment, DNA meta-barcoding, and other methods, research devoted to integrating PCC methods is needed to determine coherent approaches for characterizing PCC across scales of space, time, and diversity.

Here we assess the similarities and differences in HPLC pigment and DNA meta-barcoding PCC and PG determinations using a large data set of concurrent observations from a productive coastal ocean ecosystem, the Santa Barbara Channel (SBC), CA. Throughout this manuscript, we use PCC to refer to estimates of biomass proportions of various phytoplankton types, while PG is used to denote phytoplankton group biomass estimates in concentration units. Parallel analyses of PGs and PCC determined with each method are performed after imposing compositionality on HPLC pigment concentrations and transforming protistan ASV RSAs to estimate the concentration of particulate organic carbon associated with each ASV. Our results highlight the strengths, weaknesses, and assumptions inherent in determining PGs and PCC with DNA meta-barcoding and HPLC pigment analysis, and demonstrate that integrating HPLC and DNA meta-

barcoding observations can provide novel insights into the physiology and ecology of phytoplankton and other protists. Based on the results of covariation network analyses, we suggest that characterizing the microbial communities that covary with remotely sensible phytoplankton pigment communities offers a path to monitor ecosystem structure and function on regional to global scales via satellite ocean color observations.

Methods

Overview

All analyses presented here rely on HPLC and DNA meta-barcoding data collected from the SBC as part of the Plumes and Blooms (PnB) project. The SBC is a dynamic marine ecosystem at the boundary of the southern California Current System and Southern California Bight. Variability in physical, chemical, and biological oceanographic properties in the SBC is primarily modulated by the annual wind-driven upwelling cycle (Brzezinski and Washburn 2011; Henderikx Freitas et al. 2017). Diatoms dominate the accumulation of phytoplankton biomass associated with spring-time upwelling in the SBC, though most phytoplankton groups also respond positively to upwelling-induced nutrient enrichment of the surface ocean (Taylor et al., 2015; Catlett et al., 2021a). In addition to the annual diatom bloom, high concentrations of other phytoplankton, including dinoflagellates (Catlett et al., 2021a), prymnesiophytes (Goodman et al., 2012; Matson et al. 2019), and chlorophytes (Countway and Caron 2006), have been documented in and around the SBC.

PnB has conducted approximately monthly cruises on a north-to-south transect in the SBC and has done so since August, 1996. In the present analysis, we consider data

obtained from PnB cruises conducted between March, 2011 and September, 2014 and focus primarily on PnB observations of HPLC pigment concentrations and protistan DNA barcode sequences derived from amplicon sequencing of the V9 hypervariable region of the 18S rRNA gene (henceforth, 18S rDNA). We supplement our analysis with PnB oceanographic observations including mixed layer depth and particulate organic carbon concentrations. All PnB pigment and oceanographic data are publicly available (Catlett et al., 2020b). Raw sequence data are available in the National Center for Biotechnology Information's Sequence Read Archive (accession number PRJNA532583).

Ancillary observations of daily satellite ocean color estimates of photosynthetically available radiation (PAR) are also considered to assess the role of photophysiological variability (e.g., Goericke and Montoya, 1998) in contributing to uncertainty in PG and PCC determinations. Level 3, 4-km, daily PAR data were retrieved from the NASA Ocean Biology Distributed Active Archive Center (https://oceancolor.gsfc.nasa.gov/atbd/par/; Frouin and Pinker, 1995; Frouin et al., 2002). We constructed daily time series of PAR for each PnB station by computing mean PAR values from all 4-km pixels whose center was within 5 km of each PnB station. For most daily PnB station PAR values, 4-5 pixels were considered using this approach. In the analyses presented here, we consider the mean daily PAR values for each PnB station over the 14 days prior to each PnB sampling event. Median mixed layer light levels (Behrenfeld et al., 2005) were initially considered but PnB observations of the diffuse attenuation coefficient were not available for 23.3% of the 215 concurrent HPLC and DNA meta-barcoding observations considered here.

HPLC pigment and other oceanographic observations

Methods used for sampling and analysis of the oceanographic data considered here are available elsewhere (Catlett et al., 2020a; Catlett et al., 2021a). Briefly, potential temperature and salinity were measured with a Sea-Bird Electronics 911E CTD. Potential density was calculated using the Gibbs-SeaWater MATLAB toolbox (McDougall and Barker, 2011). Mixed layer depth (MLD) was calculated as the depth at which potential density was at least 0.2 kg m-3 higher than the mean potential density over the upper 10 m of the water column, and quality-controlled as described in Catlett et al. (2020b). Discrete seawater samples were collected near the sea surface (1 m nominal depth) in 5 L Niskin bottles. For analysis of particulate organic carbon (POC) concentrations, samples were filtered onto combusted Whatman GF/F filters (nominal pore size 0.7 μm) and frozen in liquid nitrogen until analysis at the UCSB Marine Science Institute Analytical Laboratory. Samples for analysis of phytoplankton pigment concentrations were collected on GF/F filters, frozen in liquid nitrogen, and analyzed with a standard HPLC method (Van Heukelem and Thomas 2001) at the NASA Goddard Space Flight Center. HPLC data pre-processing procedures are described in Catlett et al. (2021a).

Amplicon sequencing of 18S-V9 rDNA

Amplicon sequencing analysis of the V9 hypervariable region of the 18S rDNA follows methods outlined previously (Catlett et al., 2020a). Discrete seawater samples for amplicon sequencing analysis were collected in polycarbonate bottles which were stored and transported to the laboratory in a cooler until sampling particulate DNA within ~10 hr of collection. Approximately 1 L samples were vacuum-filtered through a 47 mm 1.2

μm mixed cellulose esters membrane filter. Filters were stored frozen in 5 mL cryovials in 1.8 mL sucrose lysis buffer (750 mmol L^{-1} sucrose, 20 mmol L^{-1} EDTA, 400 mmol L^{-1} NaCl, 50 mmol L^{-1} Tris-HCl; pH 8.0) at -80 ºC.

Genomic DNA was extracted following the phenol-chloroform method described in Catlett et al. (2020a). Briefly, lysis included mechanical (bead-beating and water bath incubations) and chemical (sodium dodecyl sulfate and proteinase K) methods. Genomic DNA was extracted with phenol-chloroform-isoamyl alcohol, followed immediately by two extractions with chloroform-isoamyl alcohol. DNA was precipitated, washed, and resuspended in Tris buffer before storage at -20 ºC. The V9 hypervariable region of the 18S rDNA was amplified with a one-step PCR using custom dual-indexed primers (Kozich et al. 2013) designed from the 1391F and EukB primers (Stoeck et al. 2010) following the "Standard" method evaluated in Catlett et al. (2020a). Following purification, normalization, and pooling of PCR products, sequencing was performed on a MiSeq PE150 v2 kit (Illumina) at the DNA Technologies Core of the UC Davis Genome Center. Each sequencing run included technical PCR/sequencing triplicates of a mock community consisting of 22 evenly represented full-length protistan 18S amplicons, at least one no-template control PCR, and multiple DNA extraction blanks. Controls were analyzed thoroughly in Catlett et al. (2020a). Data from samples amplified with certain index primers that were found to reduce precision in our DNA meta-barcoding workflow, and data from one sequencing run where one negative control showed signs of contamination, were discarded (Catlett et al., 2020a).

ASV determinations

We used the DADA2 method (Callahan et al. 2016; v1.14.1) to determine amplicon sequence variants (ASVs) from raw MiSeq data. Demultiplexed sequence reads were obtained from the UC Davis Genome Center, and forward and reverse reads were trimmed to 140 nt and 120 nt, respectively, filtered (maxEE = 2, truncQ = 2, maxN = 0), and denoised using the DADA algorithm. The DADA error model was parameterized for each MiSeq run using at least 10^8 bases. Paired reads were then merged, overhanging sequences were trimmed, and chimeras were removed using the "consensus" method (Callahan et al. 2016). ASVs less than 90 nt or greater than 180 nt in length (target amplicon is 120–130 nt) were discarded.

Ensemble taxonomic assignments and data pre-processing

Standard taxonomic assignment methods result in many ASVs with low confidence or unknown taxonomic annotations at ecologically meaningful taxonomic ranks (division, class, and lower; see Catlett et al., 2020a and Supp. Fig. S1 below). Due to difficulties in relying on a single taxonomic assignment method to classify ASVs as protists and/or phytoplankton, we implemented an ensemble taxonomic assignment approach with the goal of increasing the proportion of ASVs in our data set with taxonomy assigned at lower ranks (Catlett et al., 2021b). Initial taxonomic assignments were predicted with the RDP Bayesian classifier (Wang et al. 2007; henceforth, bayes), the DECIPHER idtaxa algorithm (Murali et al., 2018), and the Lowest Common Ancestor (LCA) algorithm implemented in MEGAN6 (Huson et al. 2007) that analyzes BLASTN (Altschul et al. 1990) results. The bayes and idtaxa algorithms used a bootstrap cutoff of 60% and 50%, respectively, and the LCA algorithm was implemented with default

parameters. All three algorithms were implemented against both the Protistan Ribosomal Reference (pr2; Guillou et al. 2012) database (v4.12.0) and the Silva SSU reference database (v138; Quast et al. 2012) available for the DADA2 pipeline (https://benjjneb.github.io/dada2/training.html). We thus generated six independent taxonomic assignments for each ASV, and refer to each collection of taxonomic assignments as idtax-pr2, idtax-silva, bayes-pr2, bayes-silva, LCA-pr2, and LCA-silva.

The ensembleTax R package (v1.1.1; Catlett et al., 2021b) was used to determine ensemble taxonomic assignments based on the six individual taxonomic assignment methods. Because assignments with the Silva, pr2, and LCA (MEGAN6 uses the NCBI taxonomy) methods employ different taxonomic naming and ranking conventions, we first mapped the idtax-silva, bayes-silva, LCA-pr2, and LCA-silva taxonomic assignments onto the pr2 taxonomic nomenclature. During mapping we retained prokaryotic assignments determined using the Silva reference database as the primers used here are not eukaryote-specific and pr2 does not include prokaryotic reference sequences. We considered the default collection of taxonomic synonyms included with the ensembleTax package for taxonomic names that were not initially mapped. All analyses presented here thus use the pr2 nomenclature.

Following mapping, we computed two sets of ensemble taxonomic assignments: the first was used to identify prokaryotic ASVs, while the second was used for the remainder of our analyses. All ensemble taxonomic assignments were determined by finding the highest frequency assignment across the (mapped, if necessary) individual taxonomic assignment methods, excluding non-assignments. In the event that conflicting taxonomic assignments were found at equivalent maximum frequencies across the six

individual methods, assignments predicted by the idtaxa algorithm, and (if the highest frequency assignment was not predicted by the idtaxa method) the bayes algorithm were prioritized. To identify prokaryotic ASVs, the bayes-pr2 taxonomic assignments (which by definition are always assigned kingdom *Eukaryota*) were omitted from ensemble determinations, and taxonomic assignments predicted with the Silva reference database were prioritized in the event multiple assignments were found at equivalent maximum frequencies. After discarding prokaryotic ASVs, a second set of ensemble taxonomic assignments was computed following the same procedure but considering all taxonomy predictions from all six individual methods and prioritizing those determined with pr2 over those determined with Silva. Overall, our approach for determining ensemble taxonomic assignments results in a higher proportion of ASVs with predicted taxonomy at lower ranks (particularly at the class rank; Supp. Fig. S1).

ASVs of non-protistan origin (those assigned as *Metazoa*, *Fungi*, *Streptophyta*, *Rhodophyta*, *Ulvophyceae*, or *Phaeophyceae*) and those that could not be identified as protists (not assigned to a kingdom or supergroup, or assigned as Eukaryota_XX, Opisthokonta_X, *Opisthokonta*, or *Archaeplastida* with unknown taxonomy at lower ranks) were discarded. Cumulative relative sequence abundances of ASVs with ambiguous identities were less than 16% in each sample, and less than 5% in 314 out of 345 total samples.

All analyses presented here rely on a data set comprising 13,308 protistan ASVs derived from 345 discrete seawater samples. Sequencing depth ranged from 11804 to 225911 protistan sequence reads per sample. Sequence counts of each protistan ASV were normalized to the total protistan sequence counts within each sample to determine

protistan ASV relative sequence abundances (RSAs) in each sample. Where duplicate or triplicate samples were available, mean RSA values were computed for each protistan ASV. Most analyses consider a subset of 215 samples collected near the sea surface, in which 6,568 total protistan ASVs were detected.

Classification of phytoplankton ASVs

Because DNA meta-barcoding data are compositional, failure to coherently identify phytoplankton ASVs will bias estimates of PCC. However, the growing recognition of mixotrophy as an important and widespread trophic strategy across diverse marine protists creates ambiguity in delineating phytoplankton from other protists (Mitra et al. 2016). Mitra et al. (2016) define four broad protistan functional groups based on modes of energy and nutrient acquisition: photoautotrophs, constitutive and non-constitutive mixotrophs, and heterotrophs. Photoautotrophs are photosynthetic carbon-fixers that do not ingest prey by phagocytosis, constitutive mixotrophs have an inherent capacity for photosynthesis but can also consume prey, non-constitutive mixotrophs must obtain their capacity for photoautotrophy through symbioses or the horizontal transfer of chloroplasts, and heterotrophs have no capacity for photosynthesis and rely on ingesting primary producers or their products to meet their energy demands (Mitra et al. 2016). Here we define phytoplankton as those taxonomic groups that are only thought to include photoautotrophic and/or constitutive mixotrophic representatives.

In order to identify phytoplankton ASVs in our data set we compiled a collection of taxonomic names with corresponding trophic modes following the definitions of Mitra et al. (2016) based on the information available in Adl et al. (2019). Where a trophic

mode was not clearly defined for a particular lineage in Adl et al. (2019), we considered additional published compilations of protistan trophic modes and traits (Schneider et al., 2020; Ramond et al., 2019; Dumack et al., 2019). After preliminary assignment of trophic modes, we conducted additional searches of both refereed (Burki et al., 2009; Okamoto and Inouye, 2005; Glucksman, 2011; Riisberg et al., 2009; Skovgaard et al., 2012; Chomerat and Bilien, 2014) and non-refereed (UC Santa Cruz Ocean Data Center, http://oceandatacenter.ucsc.edu/PhytoGallery/phytolist.html; AlgaeBase, Guiry and Guiry, 2021; and Wikipedia) sources that resulted in an additional 29 lineages in our data set assigned to trophic functional groups. Our collection of taxonomic names and corresponding trophic modes is found in Supporting File S1.

ASVs were assigned as phytoplankton, non-phytoplankton, or unknown based on their ensemble taxonomic assignments. ASVs assigned to lineages that only include photoautotrophs, constitutive mixotrophs, or both, were assigned as phytoplankton, while ASVs assigned to lineages comprised of heterotrophs and/or non-constitutive mixotrophs were assigned as non-phytoplankton. Where lineages are thought to contain representatives of both phytoplankton and non-phytoplankton, the ASV was assigned "unknown". Following this phytoplankton classification procedure, a majority (~60% or more) of sequence reads were unambiguously assigned as phytoplankton (or not) in all surface samples considered here (Supp. Fig. S2A).

Upon further inspection of the ASVs with unknown origins, the vast majority were annotated as unknown *Dinophyceae*, and to a lesser extent, unknown *Cryptophyta* or *Haptophyta*. All three of these lineages encompass a diverse group of organisms including photoautotrophs, constitutive and non-constitutive mixotrophs, and

heterotrophs. Discarding samples where these unidentified ASVs are found at high RSAs may bias the analysis to exclude samples in which these lineages are abundant. Instead, we assumed all remaining unknown *Dinophyceae*, *Cryptophyta*, and *Haptophyta* ASVs in our data were phytoplankton. With this assumption > 95% of sequence reads were classified as phytoplankton or not in the surface samples considered here (Supp. Fig. S2B). Inspection of the RSAs of phytoplankton ASVs across depth horizons (Supp. Figs. S3-4) showed that phytoplankton ASVs comprised a high proportion of the protist community in the upper 30 m of the water column that decreased with depth, providing qualitative validation of our phytoplankton classification approach.

Parallel analyses of concentrations and compositions

Since DNA meta-barcoding data are compositional and HPLC measures the concentrations of phytoplankton pigments, we performed parallel analyses of PCC and PG concentration estimates by transforming our DNA meta-barcoding data to concentrations and constraining our HPLC data to compositions. DNA meta-barcoding data were transformed to concentrations by scaling the RSAs of all protistan ASVs to concurrently determined particulate organic carbon (POC) concentrations, providing estimates of the POC associated with each protistan ASV in each sample. Estimates of POC associated with individual taxonomic groups are not impacted by the phytoplankton classification scheme employed here (with the exception of *Dinophyceae*, *Cryptophyta*, and *Haptophyta* ASVs; see above), though changes to the phytoplankton classification procedure will change estimates of phytoplankton RSAs. Considering both PCC

(composition) and PG (concentration) estimates thus provides two quasi-independent analyses that broadly corroborate one another.

We focus our analyses on the seven traditionally defined "diagnostic pigments," fucoxanthin (Fuco; biomarker for diatoms), peridinin (Perid; dinoflagellates), monovinyl chlorophyll b (MVChlb; chlorophytes), 19'-hexanoyloxyfucoxanthin (Hexfuco; prymnesiophytes), 19'-butanoyloxyfucoxanthin (Butfuco; pelagophytes), and zeaxanthin (Zea; cyanobacteria) (Vidussi et al. 2001; Uitz et al. 2006). Five of these pigments (Fuco, Perid, Hexfuco, MVChlb, and Zea) represent the five SBC PGs resolvable from HPLC pigment data alone, while Allo and Butfuco tend to covary strongly with MVChlb and Hexfuco, respectively, in the SBC (Catlett and Siegel, 2018). While ratios of biomarker pigments to TChla are commonly used and interpreted as PCC estimates, they are not truly compositional as the sum of the individual components is not constrained to a fixed value (e.g., ratios of biomarker pigments to TChla can be > 1). One approach for imposing compositionality on HPLC pigment data that requires relatively few assumptions about the underlying pigment data is that proposed by Vidussi et al. (2001). This approach normalizes the concentrations of seven diagnostic pigments to their summed concentration to derive PCC estimates. The sum of the seven diagnostic pigments is highly correlated with TChla in our data set, and so the biomass proportions derived from this approach are highly correlated with those computed using simple ratios of biomarker pigment concentrations to TChla (Supp. Fig. S5). A similar pattern was noted in Vidussi et al. (2001). These results show that ratios to TChla behave quantitatively as compositions here and likely in most other studies where they are

employed. We thus use biomarker pigment ratios to TChla to estimate PCC from HPLC pigment data.

Below we present parallel analyses of PCC (RSAs and ratios of biomarker pigments to TChla concentrations) and PGs (concentrations of biomarker pigments and POC associated with ASVs or taxonomic groups) determined by HPLC pigment and DNA meta-barcoding analysis. Neither HPLC pigment nor DNA meta-barcoding analysis measure the biomass of protistan or phytoplankton species or groups; however, for simplicity we use the terms "biomass" to refer to estimates of PG pigment or POC concentrations, "biomass proportion" to refer to PCC estimates, and "biomass contributions" to refer to both of the above.

Results

Quantification of dominant SBC PGs

Both HPLC and DNA meta-barcoding analysis indicated that the dominant eukaryotic SBC PGs are diatoms, dinoflagellates, prymnesiophytes, and chlorophytes (Fig. 1), in agreement with previous studies of HPLC pigment dynamics (Catlett et al., 2021a) and other studies (Countway and Caron, 2006; Goodman et al. 2012; Taylor et al. 2015) in and around the SBC. The highest median pigment concentrations and ratios to TChla were observed for Fuco, Hexfuco, and MVChlb, respectively. The dinoflagellate biomarker pigment, Perid, had the fourth highest median concentration amongst the biomarker pigments, and the fifth highest median ratio to TChla behind the putative cyanobacteria biomarker pigment, Zea. In the DNA meta-barcoding data, dinoflagellates had the highest median PCC and PG concentrations, followed closely by diatoms, and

then prymnesiophytes and Mamiellophyceae (a class of Chlorophyta). While Chlorophyta ASVs cumulatively had higher median PCC and PG (7.89% and 9.38 μg L-1) than Prymnesiophyceae ASVs (4.30% and 5.75 μg L-1), the prymnesiophytes comprised a higher proportion of the community than any single class within Chlorophyta.

Direct comparisons of biomass contributions determined with each method showed that estimates of PGs and PCC were broadly similar for two of the four dominant SBC PGs (Fig. 2). Strong correlations ($R2 > 0.3$, $p < 0.001$) were observed across estimates of diatom and chlorophyte biomass contributions despite a few outlier observations in each relationship. Weak correlations were observed for prymnesiophyte biomass contributions determined by the two methods ($R2 = 0.11$ and 0.26), though R2 values increased ~2-fold when two outliers were ignored (note these outliers are considered in all further analyses). Weak correlation ($R2 \leq 0.11$) was also observed between pigment- and amplicon-based estimates of dinoflagellate biomass contributions. Two distinct trends emerged in comparisons of dinoflagellate biomass contributions: a subset of stations showed dinoflagellate RSAs and POC were linearly related to Perid ratios to TChla and Perid concentrations, respectively, while another subset of stations showed increases in dinoflagellate RSA and POC with no concomitant increases in pigment-based estimates of dinoflagellate biomass contributions.

Sources of disagreement in PG biomass contribution estimates

Known sources of error in HPLC pigment assessments of PGs and PCC include physiological and inter- and intra-lineage variability in biomarker pigmentation (Higgins et al. 2011). We sought to evaluate the roles of these sources of error in the disagreements

in PCC and PG estimates noted above (Fig. 2), with any unexplained variability between the two methods likely explained by myriad sources of analytical uncertainty introduced by DNA meta-barcoding analysis. We performed multiple linear regression analysis on each phytoplankton type's PG and PCC residuals relative to the line of best fit determined by type II linear regression (Table 1; Supp. Table S2; Supp. Figs. S6-10; lines of best fit are shown in Fig. 2). Type II regression was used to define residuals in order to prevent biasing the line of best fit to favor one method over the other. Negative residual values represent overestimation of biomass contributions by pigment-based methods, while positive residual values indicate overestimation biomass contributions by DNA meta-barcoding (Fig. 2). All predictor and response variables were z-scored prior to regression analyses so that the magnitudes of regression coefficients can be directly compared across both predictor and response variables to assess the relative importance of each predictor in explaining variability in a particular set of residuals (Table 1; Supp. Table S2).

To assess the contribution of physiological variability in pigmentation to disagreements in PG and PCC estimates across the two methods, we considered mixed layer depth (MLD) as a correlate for phytoplankton physiological status since it is typically correlated with other drivers of phytoplankton physiological variability (temperature, recent light and nutrient availability). Satellite observations of photosynthetically available radiation (PAR) were also considered as a covariate of seasonal variations in phytoplankton photophysiological variability, but were not significant in predicting most phytoplankton type's residuals (see below). PAR and MLD were significantly, though weakly correlated in this data set ($r = -0.32$, $p < 0.001$), but in

most regression models this weak collinearity did not substantially impact the values of regression coefficients (Table 1; Supp. Table S2).

Selection of predictors to estimate the contribution of inter- and intra-lineage variability in biomarker pigment expression was based on known ambiguities in biomarker pigmentation (Jeffrey et al. 2011) and exploratory analyses (Supp. Figs. S6-9). Several *Dinophyceae* species express the diatom biomarker pigment, Fuco, rather than the typical dinoflagellate biomarker pigment Perid (Zapata et al. 2012). Diatom PCC and PG residuals were correlated with the cumulative PCC and PG residuals of *Dinophyceae* ASVs ($r = -0.58$ and -0.31, respectively; $p < 0.001$), with most of the variability explained by three *Dinophyceae* ASVs with PCCs that were correlated ($r < -0.4$, $p < 0.001$ for each ASV) with diatom PCC residuals. BLASTN searches of these putative Fuco-containing dinoflagellate ASV sequences against the NCBI nt database showed that they were perfect matches to 18S rDNA sequences derived from known Fuco-containing dinoflagellate genera (though there is intra-genus variability in biomarker pigmentation; Supp. Table S1). The PCC and PG concentrations of the putative Fuco-containing dinoflagellate ASVs were significantly correlated with *Dinophyceae* PCC and PG residuals ($r = 0.64$ and 0.69, respectively; $p < 0.001$). Similarly, two putative Perid-containing *Dinophyceae* ASVs were identified as those with PCC most strongly correlated ($r < -0.4$, $p < 0.001$) with *Dinophyceae* PCC residuals. Again, BLASTN searches of these ASV sequences against the NCBI nt database suggested these ASVs were derived from a Perid-containing genus (*Tripos*; Supp. Table S1).

Investigations of the potential for intra-lineage variability in biomarker pigmentation to contribute to error in pigment-based estimates of PGs and PCC showed

that *Chlorophyta* PCC and PG residuals varied systematically with the dominant *Chlorophyta* class within each sample (Supp. Fig. S7). We grouped chlorophyte classes into three "ecotypes" based on these systematic differences to use as predictors in multiple linear regression analysis. Ecotype 1 consisted of *Chlorophyceae*, *Chlorodendrophyceae*, *Chloropicophyceae*, and *Trebouxiophyceae* (classes whose dominance was associated with positive residual values), ecotype 2 included *Mammiellophyceae*, Prasino-Clade-9, and unknown *Chlorophyta* (classes whose dominance was associated with residuals with a distribution centered approximately around 0), and ecotype 3 included *Pyramimonadales* (associated with negative residual values). Further analysis showed the summed POC concentration and RSA of ecotype 1 exhibited a strong linear relationship with *Chlorophyta* PG and PCC residuals, while summed POC concentrations and RSAs of ecotypes 2 and 3 showed weaker systematic variations with Chlorophyta PG and PCC residuals (Supp. Fig. S8).

Systematic variation of prymnesiophyte PG and PCC residual values with the dominant *Prymnesiophyceae* order was not observed (Supp. Fig. S7). However, two outliers were identified in comparisons of *Prymnesiophyceae* biomass contributions across the two methods (shown in red in Fig. 2D, 2H). Further inspection revealed that these outlier observations were sampled from adjacent stations on a single cruise in April, 2011, and were dominated by a single *Prymnesiophyceae* ASV assigned as *Phaeocystis globosa* that was highly correlated with prymnesiophyte PCC and PG residuals (Supp. Fig. S9; $r = 0.72$ and 0.73 for PCC and PG correlations, respectively; $p < 0.001$). BLASTN searches against the NCBI nt database showed this putative *P. globosa* ASV

sequence was a perfect match to 18S rDNA sequences from several strains of *P. globosa*, corroborating the ensemble taxonomic assignment.

Table 1 shows multiple linear regression analysis statistics determined for predictions of PG and PCC residuals based on linear combinations of the presumed physiological correlates (MLD and PAR) and the various sources of inter- and intra-lineage variability in biomarker pigmentation identified above. Additional multiple linear regression results are included in Supporting Table S2 to further support the conclusions drawn here. MLD was a significant predictor of the PG residuals of all four dominant SBC phytoplankton types (Table 1). Negative MLD coefficient values in all PG residual regression models indicated that a deeper mixed layer was associated with relatively high pigment:POC ratios. MLD was insignificant in predicting the PCC residuals of diatoms and dinoflagellates. Conversely, MLD was significant in predicting the PCC residuals of chlorophytes and prymnesiophytes, though R2 values decreased 2% when MLD was excluded from these models (Supp. Table S2). PAR was a significant predictor of prymnesiophyte PG and PCC residuals, with positive coefficient values indicating a decrease in the ratio of Hexfuco to *Prymnesiophyceae* POC with increasing irradiance. PAR was also a significant predictor of diatom PCC residuals, though removing PAR from these regression models again resulted in a minor (3%) decrease in R2 values (Supp. Table S2).

Multiple linear regression models that included MLD and the summed POC concentrations of the three putative Fuco-containing dinoflagellate ASVs (Supp. Table S1) explained 34% of the variability in diatom PG residuals (Table 1). Standardized regression coefficients for the summed PG concentrations of the putative Fuco-containing

dinoflagellate ASVs were > 2-fold higher in magnitude than those determined for MLD. An R^2 value of 0.36 was found when predicting diatom PCC residuals with only the PCC of the three putative Fuco-containing dinoflagellate ASVs. For *Dinophyceae* PG residuals, 75% of the variability was explained by a regression model including MLD and the PG concentrations of both putative Fuco- and Perid-containing dinoflagellates (Table 1). Similar to diatom PG residuals, putative Fuco- and Perid-containing dinoflagellate concentrations had substantially larger standardized regression coefficient values than MLD. Regression models also fit dinoflagellate PCC residuals well, with 56% of the variability explained by models including the PCC of putative Fuco- and Perid-containing dinoflagellates (Table 1; Supp. Fig. S10).

Regression models that included the cumulative PG concentrations and PCC of each of the three *Chlorophyta* ecotypes accounted for 55% and 52% of the variability in *Chlorophyta* PG and PCC residuals, respectively (models used for PG residuals also included MLD; Table 1). Further, the PCC and PG concentrations of each *Chlorophyta* ecotype was significant in predicting *Chlorophyta* PCC and PG residuals, with coefficients following the trends expected based on Supp. Figs. S7-8. As observed for diatom and dinoflagellate PG residual analysis, the magnitude of standardized regression coefficients for the cumulative PG concentrations of each of the three *Chlorophyta* ecotypes were > 2-fold larger than that found for MLD, with regression coefficients for ecotype 1 having the highest magnitude. Both MLD and satellite PAR were also significant predictors of *Prymnesiophyceae* PG and PCC residuals (Table 1). Linear models including these two physiological correlates along with the PCC or PG

concentration of the putative *P. globosa* ASV as predictors explained 72 and 62% of the variability in prymnesiophyte PG and PCC residuals, respectively.

Covariation of phytoplankton classes and ASVs with biomarker pigment concentrations

Recent research demonstrates that specific communities of covarying microbes respond to oceanographic forcing and drive variability in carbon export to the deep ocean (Guidi et al. 2016; Lin et al., 2017; Caputi et al. 2018). Similarly, "communities" of biomarker pigments can be accurately reconstructed with bio-optical models and studies of their dynamics reveal the responses of phytoplankton communities to oceanographic and climate forcings (Catlett and Siegel, 2018; Catlett et al., 2021a). We suggest that community-oriented analyses may circumvent the uncertainty introduced by physiological and inter- and intra-lineage variability in biomarker pigmentation identified above and provide a path to coherently characterize phytoplankton communities across both HPLC pigment and DNA meta-barcoding analysis. Therefore, we performed network analysis to determine the dominant patterns of covariation amongst biomarker pigments and both phytoplankton and other protistan classes and ASVs observed via DNA meta-barcoding (Fig. 3 and Fig. 4 below).

Covariation networks were constructed and analyzed using the NetCoMi (Peschel et al., 2021; v1.0.2), igraph (Csardi and Nepusz, 2006; v1.2.6), and SPRING (Yoon et al., 2019; v1.0.4) R packages. We constructed four different covariation networks using the Semi-Parametric Rank-based approach for INference in Graphical models (SPRING) method (Yoon et al. 2019). The SPRING method relies on a novel estimate of pairwise

partial correlations amongst microbial ASVs or groups that accounts for the skewed, zero-inflated nature of these data. Network inference with the SPRING method relies on a data-driven optimization of a neighborhood selection approach (Meinshausen and Bühlmann, 2006; Yoon et al., 2019). Notably, this approach for inferring graphs assumes a sparse network, meaning the results shown here provide a conservative depiction of the communities of covarying biomarker pigments and ASVs or classes relative to standard Pearson correlation-based approaches (Yoon et al. 2019).

Two networks were constructed considering either class-aggregated or ASV POC concentrations alongside biomarker pigment concentrations, and two compositional networks were constructed considering either class-aggregated or ASV PCC and pigment ratios to TChla. Both phytoplankton and other protistan ASVs and classes were included in all networks in order to simultaneously determine the phytoplankton communities covarying with each biomarker pigment, as well as to assess the potential to draw inferences on ecosystem structure and function from biomarker pigment observations (see below). To maintain consistency with the analyses presented above, phytoplankton and other protists were treated as independent compositions, and classes that included both phytoplankton and other protists were separated (the prefix "phyto-" is used to denote phytoplankton where necessary to prevent ambiguity). Only those classes or ASVs that were found at > 1% PCC in at least one sample in their respective compositions were considered in network analyses. In total, 21 phytoplankton classes, 313 phytoplankton ASVs, 44 other protistan classes, and 511 other protistan ASVs were included in the network analysis. Pigment ratios to TChla and PCC of both phytoplankton and other protists were independently transformed (following Tipton et al. 2018) using

the modified centered log-ratio transformation described in Yoon et al. (2019) prior to compositional network construction. We focus our discussion here on networks constructed from class-aggregated and ASV POC concentrations given the difficulties associated with robust inference of covariation using compositional data (Yoon et al., 2019). Compositional networks are included in supporting information (Supp. Figs. S11-12) and broadly support the strongest associations observed in concentration-based networks.

To focus our discussion on positive covariation, all analyses presented here consider subsets of covariation networks with negative edges removed. We first consider phytoplankton subnetworks including only the seven diagnostic pigment and phytoplankton class (Fig. 3A) or ASV (Fig. 3B) concentrations, and expand our discussion to consider networks including both phytoplankton and other protistan classes and ASVs below (Figs. 4-7). Prior to removal of negative edges, 94.5% and 74.4% of edges represented positive associations in subnetworks including phytoplankton classes or ASVs, respectively, alongside biomarker pigment concentrations. Following removal of negative edges, both phytoplankton class and ASV subnetworks remained fully connected as all biomarker pigments and phytoplankton classes or ASVs were either linked directly by positive edges shared with one another, or linked indirectly through positive edges shared with common nodes.

Figure 3A shows the phytoplankton class and biomarker pigment concentration subnetwork. Nodes (individual phytoplankton pigments or classes) are colored according to community membership determined by a community detection algorithm that identifies communities from a network by maximizing within-community interactions

and minimizing inter-community interactions (Clauset et al., 2004), and the thickness of edges indicates the relative strength of association between nodes. Clauset et al. (2004) suggest modularity scores > 0.3 indicate "significant" community structure is resolvable from a network. A modularity score of 0.44 was found for the phytoplankton pigment and class subnetwork. In general, Figure 3A shows that phytoplankton biomarker pigment concentrations covary with diverse phytoplankton classes. As expected, phytoplankton classes tended to share a positive edge and community membership with their corresponding biomarker pigments (*Pelagophyceae* with Butfuco, *Mamiellophyceae* and *Pyramimonadales* with MVChlb, *Bacillariophyta* with Fuco, *Cryptophyceae* with Allo, and *Dinophyceae* with Perid). The exception to this pattern was *Prymnesiophyceae*, which were found in the same community as, but did not share a positive edge with, the prymnesiophyte biomarker pigment. *Dinophyceae* shared positive edges with both Fuco and Perid, but shared community membership only with Fuco. Interestingly, Zea, rather than the typical chlorophyte biomarker pigment MVChlb, shared a positive edge with the *Chlorophyta* classes *Trebouxiophyceae* and *Chloropicophyceae*, and was found in the same community as *Chlorodendrophyceae* and *Chlorophyceae*. These classes comprised *Chlorophyta* ecotype 1 defined in multiple linear regression analysis above, which was associated with consistent under-estimation of *Chlorophyta* biomass contributions from HPLC pigment data (Fig. 3; Table 1; Supp. Figs. S7-8).

Five communities of phytoplankton biomarker pigments and classes were identified by the community detection algorithm (Fig. 3A), all of which included at least one biomarker pigment. Associations amongst the biomarker pigments largely mirrored patterns of covariation in the SBC identified previously (Catlett and Siegel, 2018), with

Fuco and Zea each found in distinct communities from all other biomarker pigments, and strong covariation and shared community membership found between the chlorophyte and cryptophyte biomarker pigments (MVChlb and Allo), and between the prymnesiophyte and pelagophyte biomarker pigments (Hexfuco and Butfuco; Fig. 3A). Unexpectedly, Perid was found in the same community as MVChlb and Allo. Only four phytoplankton classes (*Chlorarachniophyceae*, Prasino-Clade-9, *Dictyochophyceae*, *Chrysophyceae*) neither shared an edge nor community membership with any biomarker pigment.

Figure 3B shows a chord diagram (Gu et al., 2014) illustrating the direct associations (positive edges) between phytoplankton biomarker pigments and ASVs. A diverse collection of phytoplankton ASVs co-varied with each biomarker pigment. Amongst the ASVs that were directly associated with each biomarker pigment, the largest proportion often belonged to the phytoplankton class for which the pigment is the corresponding biomarker. For example, Fuco shared positive edges with 7 *Bacillariophyta* ASVs, 2 *Dinophyceae* ASVs, and 1 *Prymnesiophyceae* ASV, while Perid shared positive edges with 8 *Dinophyceae* ASVs, 3 *Bacillariophyta* ASVs, and 1 *Pyramimonadales* ASV. Similarly, MVChlb shared positive edges with 5 *Chlorophyta* ASVs (4 of class *Mamiellophyceae* and 1 of *Pyramimonadales*), and 2 ASVs representative of *Dinophyceae* and *Dictyochophyceae*. Unexpectedly, Butfuco and Hexfuco shared positive edges with 1-3 ASVs from a diverse array of phytoplankton classes including *Bacillariophyta*, *Dinophyceae*, *Prymnesiophyceae*, *Pelagophyceae*, *Chloropicophyceae*, and *Cryptophyceae*. Nearly half (6/13) of the positive edges between

Allo and phytoplankton ASVs were associated with *Bacillariophyceae* ASVs rather than *Cryptophyceae* (for which Allo is a biomarker) ASVs.

Covariation of other protists with phytoplankton biomarker pigments

Phytoplankton communities tend to covary with distinct communities of microbes, including protistan micrograzers and parasites, which control the flow of phytoplankton production through marine ecosystems (Lima-Mendez et al. 2015; Berdjeb et al. 2018). Thus, in addition to characterizing the phytoplankton classes and ASVs that covary with phytoplankton pigment communities, we sought to assess covariation amongst phytoplankton pigment concentrations and non-phytoplanktonic classes and ASVs. Here, we expand the network analysis above to consider networks including both phytoplankton and other protistan classes or ASVs. Figure 4 shows subsets of these networks including only positive interactions amongst phytoplankton biomarker pigments and non-phytoplanktonic classes (Fig. 4A) or ASVs (Fig. 4B) that were not classified as phytoplankton. Whether or not phytoplankton classes or ASVs were removed from these networks, both class and ASV networks were fully connected, meaning all phytoplankton and other protistan ASVs and classes were linked to the concentrations of biomarker pigments either directly or indirectly.

Overall, phytoplankton biomarker pigments covaried with diverse protistan classes and ASVs (Fig. 4). Inspection of direct associations between biomarker pigments and protistan classes and ASVs revealed differences in the communities of protists that covaried most strongly with each biomarker pigment. In the network considering protistan class concentrations (Fig. 4A), Fuco shared edges with *Filosa-Thecofilosea* (a

class comprised primarily of nanoflagellates) and a class of radiolarians (*Polycystinea*). Perid, MVChlb, Zea, and Allo were all directly associated with (putatively non-phytoplanktonic) chrysophytes. While this was the only direct association found between non-phytoplankton classes and MVChlb, Zea also shared a positive edge with *Syndiniales*, and Perid and Allo were directly associated with several other non-phytoplanktonic classes (*Dinophyceae* and *Nassophorea* ciliates, and two *Ciliophora* clades, MAST-2, and *Polycystinea*, respectively). Hexfuco was also directly associated with *Polycystinea*, and along with Butfuco was linked to an unknown clade of *Telonemia*. Butfuco was also associated with a clade of radiolarians (RAD-B) and *Nassophorea*.

Application of the community detection algorithm (Clausnet et al., 2004) to networks including both phytoplankton and other protistan classes (modularity score of 0.42) revealed distinct community membership patterns relative to the communities identified when considering subnetworks that consisted only of phytoplankton classes. Five total communities were identified from the protistan class and phytoplankton biomarker pigment network, with three including at least one biomarker pigment (Fig. 4A). Fuco and Allo were the sole biomarker pigments found in two of the communities, both of which included a large number of non-phytoplanktonic classes. Fuco shared community membership with three classes of phytoplankton (*Bacillariophyta*, *Dinophyceae*, *Chlorophyceae*) and 14 classes of other protists, while Allo clustered with two phytoplankton classes (*Euglenozoa* and *Raphidophyceae*) in addition to nine other protistan classes, eight of which belonged to the division *Ciliophora*. Perid, Zea, MVChlb, Butfuco, and Hexfuco were all found in the same community with five phytoplankton classes and just four classes of other protists. In the other two communities that did not

include a biomarker pigment, 17 non-phytoplanktonic classes covaried with 11 phytoplankton classes, several of which (*Pelagophyceae*, *Chloropicophyceae*, MOCH-5, *Trebouxiophyceae*) shared a positive edge with at least one biomarker pigment.

Analysis of the network considering phytoplankton and other protistan ASV concentrations showed that each biomarker pigment was directly associated with distinct communities of non-phytoplanktonic ASVs derived from a diverse array of protistan classes (Fig. 4B). Fuco and Perid shared edges with ASVs from a similar distribution of classes, including non-phytoplanktonic *Dinophyceae*, *Filosa-Thecofilosea*, Spirotrich ciliates, and *Syndiniales*. Fuco also shared positive edges with several *Gregarinomorphea* ASVs. Both MVChlb and Butfuco covaried with relatively few (5-6) non-phytoplanktonic ASVs including putative nanoflagellate grazers (*Chrysophyceae*, MAST-3, and MAST-12), ciliate and dinoflagellate micrograzers (*Spirotrichea* and *Dinophyceae*), and putative parasites (*Gregarinomorphea* and *Syndiniales*). Both Zea and Hexfuco covaried with many non-phytoplanktonic ASVs, particularly those derived from *Syndiniales*. Finally, Allo covaried with 10 non-phytoplanktonic ASVs, including radiolarians, ciliates, and putative nanoflagellates.

When applied to the network including both phytoplankton and other protistan ASVs (modularity score of 0.26), the community detection algorithm identified four communities. Two communities included at least one biomarker pigment: one included Fuco along with 162 ASVs, and the other included the remaining six biomarker pigments and 284 ASVs. Thus, in total 446 of 824 ASVs included in the network analysis shared community membership with at least one biomarker pigment. About half of the ASVs that shared community membership with Fuco were not classified as phytoplankton

(84/162), while a higher fraction of non-phytoplankton ASVs covaried with the other six biomarker pigments (186/284).

Figure 5 shows the mean composition of phytoplankton and other protistan ASVs in the 215 surface ocean samples considered here for each of the four communities identified by the community detection algorithm. Unexpectedly, *Dinophyceae* ASVs comprised > 40% of mean PCC in three of the four communities. The Fuco community was the only community in which *Dinophyceae* was not the dominant phytoplankton class, and was instead dominated by diatom ASVs as expected. In the community that included the other biomarker pigments, larger contributions of ASVs derived from *Prymnesiophyceae*, *Chlorophyta*, and other phytoplankton lineages (including mostly putative pico- and nano-phytoplankton) were observed relative to the other three communities. One of the communities that lacked a biomarker pigment (Community 2) had similar mean composition to the community including six biomarker pigments, while the other (Community 1) had an unusual mean composition that included no *Chlorophyta* ASVs.

The mean composition of other protists also varied across the four communities (Fig. 5). In the community including Fuco, mean community composition was dominated by *Syndiniales* and *Gregarinomorphea* parasites, with notable contributions also observed for non-phytoplanktonic *Dinophyceae* and stramenopiles. Conversely, mean PCC of *Gregarinomorphea* ASVs was small in the community containing six biomarker pigments. Instead, the mean composition of non-phytoplanktonic ASVs in this community was relatively evenly distributed across diverse groups including ciliates, cercozoans, stramenopiles, non-phytoplanktonic dinoflagellates, and *Syndiniales*. Similar

to the patterns observed in mean phytoplankton composition, the mean composition of other protists in Community 2 was very similar to that observed for the community including most biomarker pigments, while the mean composition of other protists in Community 1 was dominated by non-phytoplanktonic dinoflagellates, ciliates, and *Syndiniales*. While high spatiotemporal variability in community composition is expected, the unique patterns in mean composition across the four communities defined here suggest unique protistan communities covary with distinct biomarker pigments.

Discussion

We compared HPLC pigment and DNA meta-barcoding estimates of PGs and PCC and identified inter- and intra-lineage variability in biomarker pigment expression (Figs. 1-2, Table 1, Supp. Table S2, Supp. Figs. S6-10) as the dominant source of disagreement across the two methods. Covariation network analyses confirmed the patterns of inter- and intra-lineage variability in pigment expression, and demonstrated that diverse communities of phytoplankton and other protists covary with biomarker pigment concentrations (Figs. 3-5). Below, we discuss the assumptions and limitations that must be accounted for when estimating PCC with HPLC pigment and DNA meta-barcoding analysis. We consider the biological mechanisms that lead to uncertainty in PG and PCC estimates, which in turn demonstrates that novel insights into phytoplankton physiological status can be gained by integrating HPLC pigment and DNA meta-barcoding observations. Finally, we suggest that the results of the network analyses above reveal a possible path toward monitoring ecosystem structure and function on

unprecedented spatiotemporal scales via satellite ocean color retrievals of phytoplankton pigment concentrations.

Limitations and assumptions involved in DNA meta-barcoding and HPLC pigment analysis of phytoplankton composition

Both DNA meta-barcoding and HPLC pigment analysis rely on critical assumptions to estimate PCC. The two major assumptions required to assess PCC with HPLC pigment data, as well as their impacts on the fidelity of HPLC PCC and PG estimates, were evident in our results. We identified inter- and intra-lineage variability in biomarker pigmentation as the largest source of uncertainty in HPLC pigment estimates of the four dominant phytoplankton types in the SBC (Fig. 2; Table 1; Supp. Figs. S6-10). Similar observations in other systems suggest this may be a consistent issue in the coastal ocean (Georicke and Montoya, 1998; Irigoien et al., 2004). Phytoplankton physiological status was also a significant contributing factor to disagreement between HPLC pigment and DNA meta-barcoding estimates of PGs and PCC, though our results indicate it is a less important source of uncertainty in the SBC (Fig. 2; Table 1). Phytoplankton physiology is a significant driver of variability in phytoplankton pigment concentrations in large portions of the open ocean (Behrenfeld et al., 2005; Siegel et al., 2013), though in conjunction with prior research our results suggest that in productive coastal areas inter- and intra-lineage variability in biomarker pigment expression is a more important source of variability in HPLC pigment PG and PCC estimates.

Many assumptions are also required to estimate PCC and PGs with DNA meta-barcoding. DNA meta-barcoding is often implicitly assumed to provide precise and

accurate estimates of PCC, though evaluations of analytical uncertainty in DNA meta-barcoding workflows are exceedingly difficult and rarely performed. Nonetheless, recent attempts to validate these methods suggest that DNA meta-barcoding workflows (including the workflow employed here) achieve comparable precision to HPLC pigment analysis and reasonable accuracy (Catlett et al. 2020a; Yeh et al., 2021). Our finding that inter- and intra-lineage variability in biomarker pigmentation is the dominant source of disagreement in HPLC pigment and DNA meta-barcoding estimates of PGs and PCC (Figs. 1-2; Table 1; Supp. Figs. S6-10) provides further support for the use of DNA meta-barcoding as a method with comparable accuracy and precision to other widely accepted methods for quantifying PCC.

This study highlighted an additional obstacle in DNA meta-barcoding analysis where assumptions are required to estimate PCC: defining and identifying phytoplankton. Several factors contribute to this issue. First, in light of the widespread occurrence of various forms of mixotrophy in marine protist communities (Mitra et al., 2016), a clear definition of what constitutes phytoplankton is lacking. A coherent definition of phytoplankton is particularly important when considering PCC estimates since the decision to classify an ASV as phytoplankton (or not) alters the PCC of all other phytoplankton ASVs. Here, we defined phytoplankton as protists (cyanobacteria are not detected with the DNA meta-barcoding method used here) with an inherent capacity to perform photosynthesis. However, some protists acquire the capacity to perform photosynthesis from other organisms (Mitra et al., 2016). Further, the nutritional modes of most marine microbes have not been directly observed, and many of the putative trophic functional group assignments compiled here rely on the assumption that

phylogenetic and/or taxonomic groups share common phenotypes in this respect (as noted in Adl et al. (2019)). One known example where this assumption is not satisfied is amongst the *Dinophyceae*, which include several genera with representatives spanning diverse nutritional modes (Adl et al., 2019). Difficulties assigning *Dinophyceae* ASVs to trophic functional groups necessitated the use of a crude assumption to classify unknown *Dinophyceae* ASVs as phytoplankton in the present analysis. Further analysis suggested that this assumption was valid (Supp. Table 1; Supp. Figs. S3-4; Fig. 2), but similar assumptions are likely to be required for years to come given the difficulties inherent in directly observing the vast diversity of protistan phenotypes.

Another factor that impacts the identification of phytoplankton in DNA meta-barcoding data is ASV taxonomy prediction. We implemented recently proposed ensemble taxonomic assignment methods (Catlett et al., 2021b) that increase the number of ASVs with taxonomy assigned at lower ranks to address this problem (Supp. Fig. S1). This approach was required here to distinguish phytoplankton from other protists, but likely comes at the expense of increased false positive annotations (Murali et al., 2018; Catlett et al., 2021b). While more detailed analyses generally supported the ensemble taxonomy predictions obtained here, there were also instances where ambiguities in ASV identification remained. For example, one of the putative Fuco-containing dinoflagellate ASVs identified here (sv8) was classified as unknown *Dinophyceae* with the ensemble methods used here, and upon further inspection it was found to be a perfect match to representatives from several dinoflagellate genera, only some of which are known to express Fuco (Supp. Table S1). This problem is partially attributable to the limited phylogenetic resolution of shorter amplicons like the 18S-V9 amplicon considered here,

though some studies suggest analysis of longer amplicons with enhanced phylogenetic resolution result in less accurate estimates of PCC (Bradley et al. 2016). Clearly, further research is needed to determine the optimal gene, hypervariable region(s), and taxonomic assignment procedures for assessing PCC.

Another major limitation of DNA meta-barcoding is the compositionality constraint inherent in these data, which can propagate analytical uncertainty and create difficulties in analyzing and interpreting variations in PCC (Aitchison, 1982; Gloor et al., 2017). However, "quantitative meta-barcoding" approaches reliant on scaling RSAs to that of a genetic internal standard (Lin et al., 2019) or to concurrent estimates of community abundance or biomass (e.g., Needham and Fuhrman 2016) allow analysts to estimate the absolute abundance or concentration of ASVs *in situ*. Here, we scaled estimates of protistan RSAs to concurrent observations of POC concentrations to estimate protistan ASV and group biomass. This scaling approach primarily relies on two key assumptions: that variability in POC concentrations is either driven by or directly proportional to variability in protist community biomass, and that variability in protistan 18S rDNA copy numbers scales with cell biovolume and biomass. The former is likely valid for our study site as the SBC is a highly productive coastal ecosystem dominated by larger-sized eukaryotic phytoplankton (Taylor et al., 2015; Catlett et al., 2021a), and previous observations in the SBC have shown that the range in bacterial biomass is generally ≤ 10% of the range in POC concentrations on seasonal to interannual time scales (Halewood et al., 2012). However, the unknown significance of detrital contributions to POC variability, as well as variability in the proportion of DNA-containing detritus of protistan origin, remains uncertain in most marine systems, and

differences in filter pore sizes used here for DNA meta-barcoding (1.2 μm) and POC analysis (0.7 μm nominal pore size) introduce further complications. Nonetheless, the magnitudes of mean POC concentrations associated with protistan groups observed here are similar to those observed in studies of nearby waters that rely on epifluorescence microscopy and carbon conversion factors to estimate biomass (Taylor et al. 2015; Caron et al. 2017). Similarly, the latter assumption is supported by several studies that document strong correlations between 18S rDNA copy number and cell size and biovolume (Zhu et al. 2005; Godhe et al. 2008; De Vargas et al. 2015), and between cell biovolume and carbon biomass (Menden-Deuer and Lassard, 2000), across diverse marine protist lineages. While further study should be devoted to validating this scaling approach against internal standard methods for estimating the concentrations of protistan groups and ASVs *in situ*, the agreement observed with HPLC pigment methods (after accounting for known sources of error in HPLC pigment methods) indicates that the POC-scaling of protistan RSAs provides reasonable estimates of protistan ASV and group biomass contributions in this data set and may be applicable to similarly productive coastal regions.

Finally, in addition to the assumptions discussed above, assessing differences in rank-order median biomass contributions observed here (Fig. 1) highlights the impacts of widely-documented weaknesses in each method. Dinoflagellate biomass proportions are severely under-estimated with HPLC pigment analysis in this data set due to the presence and quantitative significance of several putative Fuco-containing dinoflagellate species (Figs. 1-2; Table 1; Supp. Table S1; Zapata et al. 2012). Dinoflagellate biomass contributions may also be over-estimated by DNA meta-barcoding analysis due to their

anomalously high 18S rDNA copy number, although 18S rDNA copy number tends to scale with cell size and biovolume as noted above (Zhu et al., 2005; Godhe et al., 2008). We observed smaller differences in the rank-order of median *Prymnesiophyceae* and *Chlorophyta* biomass contributions across the two methods (Fig. 1), likely explained by systematic biases in estimates of *Prymnesiophyceae* PCC obtained with standard DNA meta-barcoding workflows that arises due to their anomalously high genome and 18S rDNA GC content (Liu et al., 2009, Catlett et al., 2020a). Despite the assumptions required and known weaknesses of both DNA meta-barcoding and HPLC pigment methods in estimating PGs and PCC, our ability to resolve and identify systematic errors in HPLC pigment methods (Table 1) clearly demonstrates the benefits of integrating PCC methods. Overall, our analyses highlight both widely acknowledged and more nuanced assumptions, strengths, and weaknesses in DNA meta-barcoding and HPLC pigment PG and PCC assessments.

Integrating HPLC pigment and DNA meta-barcoding analysis provides novel insights into phytoplankton physiology and ecology

We observed that both inter- and intra-lineage variability in biomarker pigment expression, and to a lesser extent, variability in phytoplankton physiological status, are significant sources of error in HPLC estimates of PGs and PCC in the SBC (Fig. 2; Table 1; Supp. Figs. S6-10). Our ability to resolve intra-lineage variability in pigment expression in smaller sized phytoplankton including prymnesiophytes and chlorophytes was particularly interesting since this has not been observed in most other studies that diagnose sources of variability by integrating HPLC and microscopic observations of

PGs and PCC (e.g., Irigoien et al., 2004). Here, we contextualize our observations of physiological and inter- and intra-lineage variability in biomarker pigmentation, and in turn, demonstrate that integrating HPLC pigment and DNA meta-barcoding observations provides novel insights into phytoplankton ecology and physiological status.

Unexpected inter-lineage variability in biomarker pigmentation was the primary source of uncertainty in diatom biomass estimates from HPLC pigment analyses, while intra-lineage variability contributed substantial uncertainty to the estimated biomass contributions of dinoflagellates, chlorophytes, and prymnesiophytes (Fig. 2; Table 1). Our results suggest that in the SBC, putative Fuco-containing dinoflagellates (Supp. Table S1) often lead to an over-estimation of diatom biomass contributions and an under-estimation of dinoflagellate biomass contributions with pigment-based methods (Fig. 2; Table 1; Supp. Fig. S6). Some of the putative Fuco-containing genera identified here (e.g., *Karenia*; Supp. Table S1) have not been observed previously in the SBC, although observations of adjacent waters have noted the presence and occasionally high concentrations of some *Gymnodinium* species (Cullen et al., 1982).

Intra-lineage variability in biomarker pigmentation was the dominant source of uncertainty in HPLC pigment estimates of chlorophyte and prymnesiophyte biomass contributions. Dominance by one of four classes within *Chlorophyta* (*Trebouxiophyceae*, *Chlorodendrophyceae*, *Chlorophyceae*, *Chloropicophyceae*, comprising ecotype 1 above) was associated with consistent underestimation of *Chlorophyta* biomass contributions with pigment methods (Fig. 2; Table 1; Supp. Figs. S7-8). While representatives of *Chlorophyceae* and *Chlorodendrophyceae* can be associated with low MVChlb expression relative to other *Chlorophyta* lineages (Higgins et al., 2011), recent studies

suggest that *Chloropicophyceae* species tend to exhibit MVChlb:TChla ratios comparable to the dominant classes included in ecotypes 2 and 3 defined above (Higgins et al., 2011; Lopes dos Santos et al., 2016; Lopes dos Santos et al., 2017). Interestingly, inspection of the PCC and PG distributions of each ecotype according to the month sampled (Supp. Fig. S13) showed that ecotype 1 tended to exhibit the highest PCC and PG concentrations during the summer (July and August), a time of year that is likely associated with a reduction in pigment:carbon ratios in many phytoplankton groups due to the relatively high surface irradiance, stratified water column, and low surface nutrient concentrations in the SBC (Behrenfeld et al., 2005; Henderikx-Freitas et al., 2017; Catlett et al., 2021a). Thus, the intra-*Chlorophyta* variability in MVChlb expression is likely due to a combination of biological (genetic) and physiological variability.

Investigation of the sources of error in prymnesiophyte determinations showed a single putative *Phaeocystis globosa* ASV (sv15) consistently contributed to an under-estimation of *Prymnesiophyceae* biomass contributions with HPLC pigment methods (Table 1; Supp. Fig. S10). *P. globosa* is a cosmopolitan, bloom-forming species that forms large colonies with cells that tend to be more carbon-rich than other *Phaeocystis* species (Schoemann et al., 2005). *P. globosa* is also known to express little to no Hexfuco when exposed to high irradiance and/or during the colonial phase of its life cycle, which supports the dependence of *Prymnesiophyceae* residuals on PAR (Fig. 2; Table 1; Schoemann et al., 2005). Altogether, our analyses in conjunction with the literature provide further support for our identification of *P. globosa* as an important bloom-forming prymnesiophyte in the SBC that is consistently unaccounted for by pigment-based PCC methods. This analysis also highlights the likely tendency of this

species to periodically form large aggregates in the SBC, which is supported by previous microscopy observations in the SBC (Goodman et al., 2012).

In addition to inter- and intra-lineage variability in biomarker pigmentation, MLD, an oceanographic correlate of temperature and recent light and nutrient availability (and presumably, of phytoplankton physiological status), was a significant predictor of the PG residuals of all four of the dominant SBC phytoplankton types (Table 1). The negative coefficient values determined for MLD in all PG residual analyses indicate deeper mixed layers were associated with increased pigment:POC concentration ratios in this data set. Increases in cellular pigment:carbon ratios in response to reduced irradiance and increased nutrient availability, both of which are typically associated with a deepening of the mixed layer, have been documented throughout the world's oceans (Behrenfeld et al., 2005 and references therein). Interestingly, MLD was either an insignificant predictor of (diatoms and dinoflagellates), or was only associated with marginal increases (< 0.025) in R^2 values for (prymnesiophytes and chlorophytes), PCC residual multiple linear regression models (Supp. Table S2). Similarly, satellite PAR was a significant predictor of prymnesiophyte PCC and PG residuals but was mostly insignificant in predicting most of the other PG and PCC residuals considered here (Supp. Table S2), possibly due to the heightened sensitivity of *P. globosa* Hexfuco expression to irradiance (Schoemann et al., 2005). Altogether, these results suggest that integrating DNA meta-barcoding and HPLC pigment data can provide novel insights into the physiological responses of dominant phytoplankton types to environmental stimuli.

Covariation network analysis confirmed the patterns of association amongst phytoplankton groups, ASVs, and biomarker pigments identified in multiple linear

regression analysis, and offered additional insights into the covariance of specific phytoplankton and other protists with different biomarker pigments (Figs. 3-5). The putative Fuco- and Perid-containing dinoflagellate ASVs identified above (Supp. Table S1) were either directly associated, or shared community membership, with their putative biomarker pigments. Additional phytoplanktonic dinoflagellate ASVs were directly associated with Perid, including ASVs assigned to genera known to express Perid (e.g., *Alexandrium*, *Heterocapsa*; Zapata et al., 2012). Interestingly, one putative Perid-containing ASV (assigned to the Perid-containing genus *Paragymnodinium*; Kang et al., 2010) was directly linked to both Fuco and Perid, pointing to the difficulties in identifying the biomarker pigments of different phytoplankton species *in situ*. Network analysis also confirmed that both *Mamiellophyceae* and *Pyramimonadales* (*Chlorophyta* ecotypes 2 and 3) covary with MVChlb while *Chlorophyta* classes included in ecotype 1 covary more strongly with Zea (Fig. 3A). Notably, the single *Prymnesiophyceae* ASV that was directly associated with Fuco (Fig. 3B) was the same *P. globosa* ASV (sv15) that was identified as the primary source of variability in *Prymnesiophyceae* biomarker pigmentation. *P. globosa* often continues to express Fuco as Hexfuco expression is reduced, providing support for this association (Schoemann et al., 2005).

Finally, one particularly interesting example of a novel perspective of the microbial dynamics contributing to biomarker pigment variations was found upon closer inspection of the protistan ASVs that shared positive edges with the crypophyte biomarker pigment, Allo, in network analyses (Fig. 4B). Allo was associated with two ASVs classified as the ciliate, *Mesodinium sp.*, and three cryptophyte ASVs classified as *Teleaulax*. Figure 6 shows this network motif alongside the spatiotemporal dynamics of

Allo and associated *Mesodinium* and *Teleaulax* ASVs. *Mesodinium* includes species with diverse nutritional strategies (Garcia-Cuetos et al., 2012). Representatives of the *M. rubrum* species complex can acquire the capacity for photoautotrophy via ingestion of *Teleaulax* and subsequently preserve and maintain transcriptionally active *Teleaulax* chloroplasts and nuclei (Qiu et al., 2016). The covariation amongst *Mesodinium* and *Teleaulax* ASVs and Allo suggests that these two *Mesodinium* ASVs are derived from *M. rubrum*, and that this organism regularly relies on *Teleaulax* prey and acquired photoautotrophy to meet its nutritional requirements in the SBC. Further, the spatiotemporal dynamics of all of the above demonstrates that the putative *M. rubrum* ASVs are associated with the highest concentrations of Allo observed in this data set. Periods with relatively high concentrations of particular *Teleaulax* ASVs in conjunction with low concentrations of *Mesodinium*, and vice-versa, may indicate periods where *Teleaulax* escaped predation, or where *Mesodinium* was reliant on other nutritional strategies (Fig. 6). While detailed analyses of these population dynamics are beyond the scope of this work, this example in conjunction with the analysis of physiological and inter- and intra-lineage variability in biomarker pigment expression above provides strong evidence that integrating HPLC pigment and DNA meta-barcoding data provides novel insights into the physiological status and ecology of phytoplankton and other protists.

Implications for satellite remote sensing: toward assessing ecosystem structure and function from ocean color?

Perhaps the greatest strength in HPLC pigment assessments of PCC is that pigments are bio-optically active and have unique spectral absorption properties (Bricaud et al., 2004; Catlett and Siegel, 2018). These relationships provide the motivation for recent attempts to retrieve phytoplankton biomarker pigment concentrations and/or derived PCC indices from remotely sensible bio-optical properties (Uitz et al., 2015; Chase et al., 2017). The imminent launch of the NASA Plankton, Aerosols, Clouds, and ocean Ecosystems (PACE) mission (Werdell et al., 2019) is expected to improve retrievals of biomarker pigment concentrations from satellite ocean color observations by improving resolution of small-scale phytoplankton absorption features (Uitz et al., 2015; Catlett and Siegel, 2018), thus offering a path to observe PCC on unprecedented spatiotemporal scales.

Recent work has identified "communities" of phytoplankton pigments that covary with one another on both regional (Latasa and Bidigare, 1998; Catlett and Siegel, 2018; Kramer et al., 2020) and global (Kramer and Siegel, 2018) scales. The covariation network analysis applied here confirms previously documented patterns of covariation amongst phytoplankton biomarker pigment concentrations in the SBC, and demonstrates that diverse communities of both phytoplankton and other protistan classes and ASVs are associated with these phytoplankton pigment communities (Figs. 3-6; Supp. Figs. S11-12). It follows that unique suites of food web interactions amongst phytoplankton and other protists covary with distinct (suites of) biomarker pigments. As a preliminary test of this hypothesis, we inspected the interactions between phytoplankton and other protistan ASVs found within each of the four communities identified by the community detection algorithm (Clauset et al., 2004) when applied to the protistan ASV POC and biomarker

pigment concentration covariation network constructed above (see Fig. 5 for mean composition of these communities).

Figure 7 shows the fraction of positive interactions (edges) between phytoplankton and other protistan ASVs (interpreted here as putative food web interactions) found in each of the four communities. Each community displayed unique collections of putative food web interactions. For example, in the community that included Fuco, diatoms accounted for ~60% of putative food web interactions, while chlorophytes, prymnesiophytes, and other phytoplankton cumulatively comprised < 10% of food web interactions. Surprisingly, approximately half of the interactions between diatoms and non-phytoplankton in the community containing Fuco were with putative nanoflagellate grazers, parasites, and parasitoids (*Syndiniales*, *Gregarinomorphea*, *Cercozoa*, *Stramenopiles*), suggesting more detailed analyses may help refine and improve size-structured food web models, which in turn will allow for more robust predictions of marine ecosystem function (e.g., Ward et al., 2012; Siegel et al., 2014). Conversely, in the community containing the other six biomarker pigments, phytoplanktonic dinoflagellates and ciliates accounted for a larger proportion of putative food web interactions. While a complete analysis of these interactions is beyond the scope of the present analysis, these broad differences suggest that phytoplankton pigment communities are representative of distinct food web dynamics.

Despite the potential to remotely sense marine ecosystem structure and function from satellite ocean color demonstrated by our analyses, several major limitations and questions remain. First, identifying (partial) correlation is not equivalent to directly verifying and identifying the type, frequency, and efficiency of interactions amongst

organisms (Fuhrman et al., 2015). Further, network analysis and other methods for assessing interaction and covariation amongst microbial communities are developing rapidly (Gloor et al., 2017; Yoon et al., 2019), and the optimal methods for characterizing microbial covariability in pelagic ecosystems remain largely unknown. Covariation-based analyses are inherently empirical, and patterns of covariation and microbial association may vary across the relevant spatiotemporal scales for understanding and predicting marine ecosystem function (Fuhrman et al., 2015). Our analysis clearly illustrates the empirical nature of these approaches, as community membership determined by the community detection algorithm employed here depends on network topology (Clauset et al., 2004; note the variability in biomarker pigment community membership in Figs. 3A and 4A), which in turn depends on the features (ASVs, classes, and/or biomarker pigments) included in network analysis. Further, the predictability of interactions that covary indirectly with biomarker pigments, which comprise the majority of putative food web interactions observed here, remains largely unknown. Employing these approaches to inform large scale data-driven food web models will thus require more detailed study of the spatiotemporal variability in and predictive relationships amongst biomass contributions and the frequencies and efficiencies of interactions amongst microbial species and groups, in addition to broadly characterizing the identities and mean composition of interacting microbes as done here (Figs. 5, 7).

Nonetheless, characterizing the microbial communities and interactions that covary with biomarker pigment concentrations has several benefits relative to other approaches to assess PCC from satellite ocean color. First, this approach requires minimal *a priori* assumptions and circumvents the two major obstacles in quantifying

PCC from HPLC pigment observations: inter- and intra-lineage variability in biomarker pigment expression and phytoplankton physiology. Second, covariation-based approaches can be applied to estimate PCC at any temporal and/or spatial scale, as well as across any "scale" of microbial diversity. The latter point is clearly demonstrated in our analyses, which show that diverse protistan assemblages at both the class and ASV levels covary with distinct suites of phytoplankton biomarker pigments (Figs. 3-7). Given the potential importance of microdiversity in determining the roles of microbes in marine ecosystems (e.g., Needham et al., 2017; Treguer et al., 2018), the potential to gain insight into microbial interactions across scales of plankton diversity may be particularly important in improving our ability to predict large-scale ecosystem dynamics. The potential to consistently characterize PCC across various spatiotemporal scales contrasts with other approaches for estimating PCC from HPLC pigments that require assumptions of fixed biomarker pigment ratios to TChla and/or negligible covariation amongst phytoplankton groups across all scales of space and time, and that derive biomass contributions of phytoplankton "pigment types" whose phylogenetic, ecological and geochemical relevance remains unclear (Higgins et al., 2011). Although the optimal observational scales of time, space, and plankton diversity for relating PCC to ecosystem structure and function remain to be determined, community-oriented analyses offer a path to address these questions. Overall, our analysis shows that characterizing the microbial communities and interactions that covary with biomarker pigment concentrations provides a potential path to monitor marine ecosystem structure and function from ocean color remote sensing.

Acknowledgements

We thank the Plumes and Blooms team, especially Nathalie Guillocheau and Stuart Halewood, for their efforts collecting the data and samples analyzed here. We also thank Jeff Krause for sharing the DNA meta-barcoding samples analyzed here. Plumes and Blooms is funded by the National Aeronautics and Space Administration (NASA; NNX11AL94G), and ship time for the data presented here was provided by NASA and the National Oceanic and Atmospheric Administration (NOAA) Channel Islands National Marine Sanctuary. Additional support of this work and of DC was provided by the NASA Biodiversity and Ecological Forecasting program (Grant NNX14AR62A), the Bureau of Ocean and Energy Management Ecosystem Studies program (BOEM award MC15AC00006) and NOAA in support of the Santa Barbara Channel Marine Biodiversity Observation Network, and by the NASA PACE Science Team (Grant 80NSSC20M0226). DC was also funded in part by a NASA Earth and Space Science Fellowship (Grant NNX16AO44HS02).

References

Adl, S. M., D. Bass, C. E. Lane, and others. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. Journal of Eukaryotic Microbiology **66**: 4–119.

Aitchison, J. (1982). The statistical analysis of compositional data. Journal of the Royal Statistical Society: Series B (Methodological), 44(2), 139-160.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of molecular biology **215**: 403–410.

Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global biogeochemical cycles*, *19*(1). https://doi.org/10.1029/2004GB002299.

Berdjeb, L., A. Parada, D. M. Needham, and J. A. Fuhrman. 2018. Short-term dynamics and interactions of marine protist communities during the spring–summer transition. The ISME journal **12**: 1907.

Bradley, I. M., A. J. Pinto, and J. S. Guest. 2016. Design and evaluation of Illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. Appl. Environ. Microbiol. **82**: 5878–5891.

Bricaud, A., H. Claustre, J. Ras, and K. Oubelkheir (2004), Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations, *Journal of Geophysical Research-Oceans*, *109*(C11), doi:10.1029/2004jc002419.

Brzezinski, M. A., and L. Washburn. 2011. Phytoplankton primary productivity in the Santa Barbara Channel: Effects of wind-driven upwelling and mesoscale eddies. Journal of Geophysical Research: Oceans **116**.

Burki, F., Inagaki, Y., Bråte, J., Archibald, J.M., Keeling, P.J., Cavalier-Smith, T., Sakaguchi, M., Hashimoto, T., Horak, A., Kumar, S. and Klaveness, D., 2009. Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, Telonemia and Centroheliozoa, are related to photosynthetic chromalveolates. *Genome biology and evolution*, *1*, pp.231-238. https://doi.org/10.1093/gbe/evp022.

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

Caputi, L., Q. Carradec, D. Eveillard, and others. 2018. Community-Level Responses to Iron Availability in Open Ocean Planktonic Ecosystems. Global Biogeochemical Cycles.

Caron, D. A., P. E. Connell, R. A. Schaffner, A. Schnetzer, J. A. Fuhrman, P. D. Countway, and D. Y. Kim. 2017. Planktonic food web structure at a coastal time-series site: I. Partitioning of microbial abundances and carbon biomass. Deep Sea Research Part I: Oceanographic Research Papers **121**: 14–29.

Catlett, D., and D. A. Siegel. 2018. Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment. Journal of Geophysical Research: Oceans **123**: 246–264.

Catlett, D., Matson, P. G., Carlson, C. A., Wilbanks, E. G., Siegel, D. A., & Iglesias-Rodriguez, M. D. (2020a). Evaluation of accuracy and precision in an amplicon sequencing workflow for marine protist communities. *Limnology and Oceanography: Methods*, *18*(1), 20-40.

Catlett D, Siegel DA, Guillocheau N. 2020b. Plumes and Blooms: Curated oceanographic and phytoplankton pigment observations ver 1. Environmental Data Initiative. https://doi.org/10.6073/pasta/f88ee1dc32b8785fe6ce57d80722e78c.

Catlett, D., Siegel, D. A., Simons, R. D., Guillocheau, N., Henderikx-Freitas, F., & Thomas, C. S. (2021a). Diagnosing seasonal to multi-decadal phytoplankton group dynamics in a highly productive coastal ecosystem. *Progress in Oceanography*, 102637. https://doi.org/10.1016/j.pocean.2021.102637.

Catlett D, Son K, Liang C. 2021b. ensembleTax: an R package for determinations of ensemble taxonomic assignments of phylogenetically-informative marker gene sequences. *PeerJ* 9:e11865 https://doi.org/10.7717/peerj.11865.

Chase, A., E. Boss, I. Cetinić, and W. Slade. 2017. Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra: toward a global algorithm. Journal of Geophysical Research: Oceans **122**: 9725–9743.

Chomerat, N., & Bilien, G. (2014). Madanidinium loirii gen. et sp. nov.(Dinophyceae), a new marine benthic dinoflagellate from Martinique Island, Eastern Caribbean. *European Journal of Phycology*, *49*(2), 165-178. https://doi.org/10.1080/09670262.2014.898797

Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, *70*(6), 066111. https://doi.org/10.1103/PhysRevE.70.066111.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, complex systems*, *1695*(5), 1-9.

Countway, P. D., and D. A. Caron. 2006. Abundance and distribution of Ostreococcus sp. in the San Pedro Channel, California, as revealed by quantitative PCR. Appl. Environ. Microbiol. **72**: 2496–2506.

Cullen, J. J., Horrigan, S. G., Huntley, M. E., & Reid, F. M. (1982). Yellow water in La Jolla Bay, California, July 1980. I. A bloom of the dinoflagellate, Gymnodinium flavum Kofoid & Swezy. *Journal of experimental marine biology and ecology*, *63*(1), 67-80. https://doi.org/10.1016/0022-0981(82)90051-X.

De Vargas, C., S. Audic, N. Henry, and others. 2015. Eukaryotic plankton diversity in the sunlit ocean. Science **348**: 1261605.

Dumack, K., Fiore-Donno, A. M., Bass, D., & Bonkowski, M. (2020). Making sense of environmental sequencing data: ecologically important functional traits of the protistan groups Cercozoa and Endomyxa (Rhizaria). *Molecular ecology resources*, *20*(2), 398-403. https://doi.org/10.1111/1755-0998.13112.

Frouin, R. & Pinker, R. T. (1995). Estimating Photosynthetically Active Radiation (PAR) at the earth's surface from satellite observations. Remote Sensing of Environment, Volume 51, Issue 1, January 1995, Pages 98-107, ISSN 0034-4257. http://dx.doi.org/10.1016/0034-4257(94)00068-X.

Frouin, R., Franz, B. A., & Werdell, P. J. (2002). The SeaWiFS PAR product. ,In: S.B. Hooker and E.R. Firestone, Algorithm Updates for the Fourth SeaWiFS Data Reprocessing, NASA Tech. Memo. 2003-206892, Volume 22, NASA Goddard Space Flight Center, Greenbelt, Maryland, 46-50.

Fuhrman, J. A., Cram, J. A., & Needham, D. M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology*, *13*(3), 133-146.

Garcia-Cuetos, L., Moestrup, Ø., & Hansen, P. J. (2012). Studies on the genus Mesodinium II. Ultrastructural and molecular investigations of five marine species help clarifying the taxonomy. *Journal of Eukaryotic Microbiology*, *59*(4), 374-400. https://doi.org/10.1111/j.1550-7408.2012.00630.x.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. *Frontiers in microbiology*, *8*, 2224. https://doi.org/10.3389/fmicb.2017.02224.

Glücksman, E., 2011. *Taxonomy, biodiversity, and ecology of Apusozoa (Protozoa)* (Doctoral dissertation, Oxford University, UK).

Godhe, A., M. E. Asplund, K. Härnström, V. Saravanan, A. Tyagi, and I. Karunasagar. 2008. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. Appl. Environ. Microbiol. **74**: 7174–7182.

Goericke, R., & Montoya, J. P. (1998). Estimating the contribution of microalgal taxa to chlorophyll a in the field--variations of pigment ratios under nutrient-and light-limited growth. *Marine Ecology Progress Series*, *169*, 97-112. doi:10.3354/meps169097.

Gong, W., Hall, N., Paerl, H., & Marchetti, A. (2020). Phytoplankton composition in a eutrophic estuary: Comparison of multiple taxonomic approaches and influence of environmental factors. *Environmental Microbiology*, *22*(11), 4718-4731. https://doi.org/10.1111/1462-2920.15221.

Goodman, J., M. A. Brzezinski, E. R. Halewood, and C. A. Carlson. 2012. Sources of phytoplankton to the inner continental shelf in the Santa Barbara Channel inferred from cross-shelf gradients in biological, physical and chemical parameters. Continental Shelf Research **48**: 27–39.

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics*, *30*(19), 2811-2812. https://doi.org/10.1093/bioinformatics/btu393.

Guidi, L., S. Chaffron, L. Bittner, and others. 2016. Plankton networks driving carbon export in the oligotrophic ocean. Nature **532**: 465.

Guillou, L., D. Bachar, S. Audic, and others. 2012. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic acids research **41**: D597–D604.

Guiry, M.D. in Guiry, M.D. & Guiry, G.M. 2021. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. http://www.algaebase.org; searched on 14 June 2021.

Henderikx Freitas, F., D. A. Siegel, S. Maritorena, and E. Fields. 2017. Satellite assessment of particulate matter and phytoplankton variations in the S anta B arbara C hannel and its surrounding waters: R ole of surface waves. Journal of Geophysical Research: Oceans **122**: 355–371.

Higgins, H. W., S. W. Wright, and L. Schluter. 2011. Quantitative interpretation of chemotaxonomic pigment data.

Huson, D. H., A. F. Auch, J. Qi, and S. C. Schuster. 2007. MEGAN analysis of metagenomic data. Genome research **17**: 377–386.

Irigoien, X., Meyer, B., Harris, R., & Harbour, D. (2004). Using HPLC pigment analysis to investigate phytoplankton taxonomy: the importance of knowing your species. *Helgoland Marine Research*, *58*(2), 77-82. DOI 10.1007/s10152-004-0171-9.

Jeffrey, S. W., S. W. Wright, and M. Zapata. 2011. Microalgal classes and their signature pigments.

Kozich, J. J., S. L. Westcott, N. T. Baxter, S. K. Highlander, and P. D. Schloss. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. **79**: 5112–5120.

Kramer, S., and D. Siegel. 2018. Global surface ocean phytoplankton community structure determined from co-variability in phytoplankton pigment concentrations.

Kramer, S. J., Siegel, D. A., & Graff, J. R. (2020). Phytoplankton community composition determined from co-variability among phytoplankton pigments from the NAAMES field campaign. *Frontiers in Marine Science*, *7*, 215. https://doi.org/10.3389/fmars.2020.00215.

Le Quere, C., et al. (2005), Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, Global Change Biology, 11(11), 2016-2040, doi:10.1111/j.1365-2468.2005.01004.x.

Lima-Mendez, G., K. Faust, N. Henry, and others. 2015. Determinants of community structure in the global plankton interactome. Science **348**: 1262073.

Lin, Y., Gifford, S., Ducklow, H., Schofield, O., & Cassar, N. (2019). Towards quantitative microbiome community profiling using internal standards. *Applied and environmental microbiology*, *85*(5), e02634-18. https://doi.org/10.1128/AEM.02634-18.

Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F. and de Vargas, C., 2009. Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proceedings of the national academy of sciences*, *106*(31), pp.12803-12808. https://doi.org/10.1073/pnas.0905841106.

Lombard, F., E. Boss, A. M. Waite, and others. 2019. Globally consistent quantitative observations of planktonic ecosystems. Frontiers in Marine Science **6**: 196.

Lopes dos Santos, A., Gourvil, P., Rodríguez, F., Garrido, J. L., & Vaulot, D. (2016). Photosynthetic pigments of oceanic Chlorophyta belonging to prasinophytes clade VII. Journal of phycology, 52(1), 148-155.

Lopes dos Santos, A., Pollina, T., Gourvil, P., Corre, E., Marie, D., Garrido, J.L., Rodríguez, F., Noël, M.H., Vaulot, D. and Eikrem, W., 2017. Chloropicophyceae, a new class of picophytoplanktonic prasinophytes. *Scientific reports*, *7*(1), pp.1-20.

Mackey, M., D. Mackey, H. Higgins, and S. Wright. 1996. CHEMTAX-a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. Marine Ecology Progress Series **144**: 265–283.

Matson, P. G., L. Washburn, E. A. Fields, C. Gotschalk, T. M. Ladd, D. A. Siegel, Z. S. Welch, and M. D. Iglesias-Rodriguez. 2019. Formation, development, and propagation of a rare coastal coccolithophore bloom. Journal of Geophysical Research: Oceans **124**: 3298–3316.

McDougall, T.J. and P.M. Barker, 2011: Getting started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox, 28pp., SCOR/IAPSO WG127, ISBN 978-0-646-55621-5.

Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, *34*(3), 1436-1462. DOI: 10.1214/009053606000000281.

Menden-Deuer, S., & Lessard, E. J. (2000). Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnology and oceanography*, *45*(3), 569-579. https://doi.org/10.4319/lo.2000.45.3.0569.

Mitra, A., Flynn, K.J., Tillmann, U., Raven, J.A., Caron, D., Stoecker, D.K., Not, F., Hansen, P.J., Hallegraeff, G., Sanders, R. and Wilken, S., 2016. Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: incorporation of diverse mixotrophic strategies. *Protist*, *167*(2), pp.106-120.

Murali, A., Bhargava, A., & Wright, E. S. (2018). IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome, 6(1), 1-14. https://doi.org/10.1186/s40168-018-0521-5.

Needham, D. M., and J. A. Fuhrman. 2016. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nature Microbiology **1**: 16005.

Needham, D. M., Sachdeva, R., & Fuhrman, J. A. (2017). Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *The ISME journal*, *11*(7), 1614-1629. doi:10.1038/ismej.2017.29.

Okamoto, N., & Inouye, I. (2005). The katablepharids are a distant sister group of the Cryptophyta: a proposal for Katablepharidophyta divisio nova/Kathablepharida phylum novum based on SSU rDNA and beta-tubulin phylogeny. *Protist*, *156*(2), 163-179. https://doi.org/10.1016/j.protis.2004.12.003.

Peschel, S., Müller, C. L., von Mutius, E., Boulesteix, A. L., & Depner, M. (2021). NetCoMi: network construction and comparison for microbiome data in R. *Briefings in Bioinformatics*, *22*(4), bbaa290. https://doi.org/10.1093/bib/bbaa290.

Qiu, D., Huang, L., & Lin, S. (2016). Cryptophyte farming by symbiotic ciliate host detected in situ. *Proceedings of the National Academy of Sciences*, *113*(43), 12208-12213. https://doi.org/10.1073/pnas.1612483113.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research **41**: D590–D596.

Ramond, P., Sourisseau, M., Simon, N., Romac, S., Schmitt, S., Rigaut-Jalabert, F., Henry, N., De Vargas, C. and Siano, R., 2019. Coupling between taxonomic and

functional diversity in protistan coastal communities. *Environmental microbiology*, *21*(2), pp.730-749. https://doi.org/10.1111/1462-2920.14537.

Riisberg, I., Orr, R.J., Kluge, R., Shalchian-Tabrizi, K., Bowers, H.A., Patil, V., Edvardsen, B. and Jakobsen, K.S., 2009. Seven gene phylogeny of heterokonts. *Protist*, *160*(2), pp.191-204. https://doi.org/10.1016/j.protis.2008.11.004.

Ryther, J. H. 1969. Photosynthesis and fish production in the sea. Science **166**: 72–76.

Schlüter, L., Møhlenberg, F., Havskum, H., & Larsen, S. (2000). The use of phytoplankton pigments for identifying and quantifying phytoplankton groups in coastal areas: testing the influence of light and nutrients on pigment/chlorophyll a ratios. *Marine Ecology progress series*, *192*, 49-63.

Schneider, L.K., Anestis, K., Mansour, J., Anschütz, A.A., Gypens, N., Hansen, P.J., John, U., Klemm, K., Martin, J.L., Medic, N. and Not, F., 2020. A dataset on trophic modes of aquatic protists. Biodiversity data journal, 8. doi: 10.3897/BDJ.8.e56648

Schoemann, V., Becquevort, S., Stefels, J., Rousseau, V., & Lancelot, C. (2005). Phaeocystis blooms in the global ocean and their controlling mechanisms: a review. *Journal of Sea Research*, *53*(1-2), 43-66. https://doi.org/10.1016/j.seares.2004.01.008.

Siegel, D. A., Buesseler, K. O., Doney, S. C., Sailley, S. F., Behrenfeld, M. J., & Boyd, P. W. (2014). Global assessment of ocean carbon export by combining satellite observations and food-web models. *Global Biogeochemical Cycles*, *28*(3), 181-196.

Skovgaard, A., Karpov, S. A., & Guillou, L. (2012). The parasitic dinoflagellates Blastodinium spp. inhabiting the gut of marine, planktonic copepods: morphology, ecology, and unrecognized species diversity. *Frontiers in microbiology*, *3*, 305. https://doi.org/10.3389/fmicb.2012.00305.

Stoeck, T., D. Bass, M. Nebel, R. Christen, M. D. Jones, H. BREINER, and T. A. Richards. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular ecology **19**: 21–31.

Taylor, A. G., M. R. Landry, K. E. Selph, and J. J. Wokuluk. 2015. Temporal and spatial patterns of microbial community biomass and composition in the Southern California Current Ecosystem. Deep Sea Research Part II: Topical Studies in Oceanography **112**: 117–128.

Tipton, L., Müller, C.L., Kurtz, Z.D., Huang, L., Kleerup, E., Morris, A., Bonneau, R. and Ghedin, E., 2018. Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome*, *6*(1), pp.1-14. https://doi.org/10.1186/s40168-017-0393-0.

Trudnowska, E., Lacour, L., Ardyna, M. *et al.* Marine snow morphology illuminates the evolution of phytoplankton blooms and determines their subsequent vertical export. *Nat Commun* **12,** 2816 (2021). https://doi.org/10.1038/s41467-021-22994-4

Uitz, J., H. Claustre, A. Morel, and S. B. Hooker. 2006. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. Journal of Geophysical Research: Oceans **111**.

Van Heukelem, L., and C. S. Thomas. 2001. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. Journal of Chromatography A **910**: 31–49.

Vidussi, F., H. Claustre, B. B. Manca, A. Luchetta, and J. Marty. 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. Journal of Geophysical Research: Oceans **106**: 19939–19956.

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. **73**: 5261–5267.

Wear, E. K., Wilbanks, E. G., Nelson, C. E., & Carlson, C. A. (2018). Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton. *Environmental microbiology*, *20*(8), 2709-2726. https://doi.org/10.1111/1462-2920.14091.

Werdell, P.J., Behrenfeld, M.J., Bontempi, P.S., Boss, E., Cairns, B., Davis, G.T., Franz, B.A., Gliese, U.B., Gorman, E.T., Hasekamp, O. and Knobelspiesse, K.D., 2019. The Plankton, Aerosol, Cloud, ocean Ecosystem mission: status, science, advances. *Bulletin of the American Meteorological Society*, *100*(9), pp.1775-1794. https://doi.org/10.1175/BAMS-D-18-0056.1.

Yeh, Y. C., McNichol, J., Needham, D. M., Fichot, E. B., Berdjeb, L., & Fuhrman, J. A. (2021). Comprehensive single-PCR 16S and 18S rRNA community analysis

validated with mock communities, and estimation of sequencing bias against 18S. *Environmental Microbiology*. https://doi.org/10.1111/1462-2920.15553.

Yoon, G., Gaynanova, I., & Müller, C. L. (2019). Microbial networks in SPRING-Semi-parametric rank-based correlation and partial correlation estimation for quantitative microbiome data. *Frontiers in genetics*, *10*, 516. https://doi.org/10.3389/fgene.2019.00516.

Zapata, M., S. Fraga, F. Rodríguez, and J. L. Garrido. 2012. Pigment-based chloroplast types in dinoflagellates. Marine Ecology Progress Series **465**: 33–52.

Zhu, F., R. Massana, F. Not, D. Marie, and D. Vaulot. 2005. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS microbiology ecology **52**: 79–92.

Table 1. Multiple linear regression analysis of type II regression residuals (see Fig. 2) for the PG and PCC estimates of the four dominant SBC phytoplankton types. All variables were z-scored prior to multiple regression analysis. Coefficients of determination (R^2) and sample size (N) are shown for each multiple regression model. Standardized regression coefficients (p-values) are shown for each predictor included in multiple regression analysis. Predictor variable abbreviations: MLD, mixed layer depth; PAR, photosynthetically available radiation; Fuco-dino, summed POC or RSA of three putative Fuco-containing dinoflagellate ASVs (see main text, Supp. Table S1); Perid-dino, summed POC or RSA of three putative Perid-containing dinoflagellate ASVs (see main text, Supp. Table S1); Chloro-eco1, Chloro-eco2, Chloro-eco3, summed POC or RSA of all ASVs included in *Chlorophyta* ecotypes 1, 2, or 3, respectively (see main text, Supp. Figs. S7-8); sv15-*P. globosa*, the POC or RSA of the putative *Phaeocystis globosa* ASV associated with underestimation of prymnesiophyte biomass contributions with HPLC pigment methods (see main text, Supp. Fig. S9).

		R^2	N	MLD (p-val)	PAR (p-val)	Fuco-dino* (p-val)	Perid-dino* (p-val)	Chloro-eco1* (p-val)	Chloro-eco2* (p-val)	Chloro-eco3* (p-val)	sv15-*P. globosa** (p-val)
Diatom	PG	0.34	183	-0.23 (< 0.001)		-0.51 (< 0.001)					
	PCC	0.36	189			-0.60 (< 0.001)					
Dino	PG	0.75	183	-0.16 (0.001)		0.72 (< 0.001)	-0.39 (< 0.001)				
	PCC	0.56	189			0.65 (< 0.001)	-0.28 (< 0.001)				
Chloro	PG	0.55	183	-0.16 (0.002)				0.60 (< 0.001)	0.34 (< 0.001)	-0.39 (< 0.001)	
	PCC	0.52	189					0.69 (< 0.001)	0.20 (< 0.001)	-0.20 (< 0.001)	
Prym	PG	0.62	183	-0.12 (0.02)	0.28 (< 0.001)						0.70 (< 0.001)
	PCC	0.72	189		0.39 (< 0.001)						0.73 (< 0.001)

*POC concentrations were used to predict PG POC residuals, while RSAs were used to predict PCC RSA residuals.

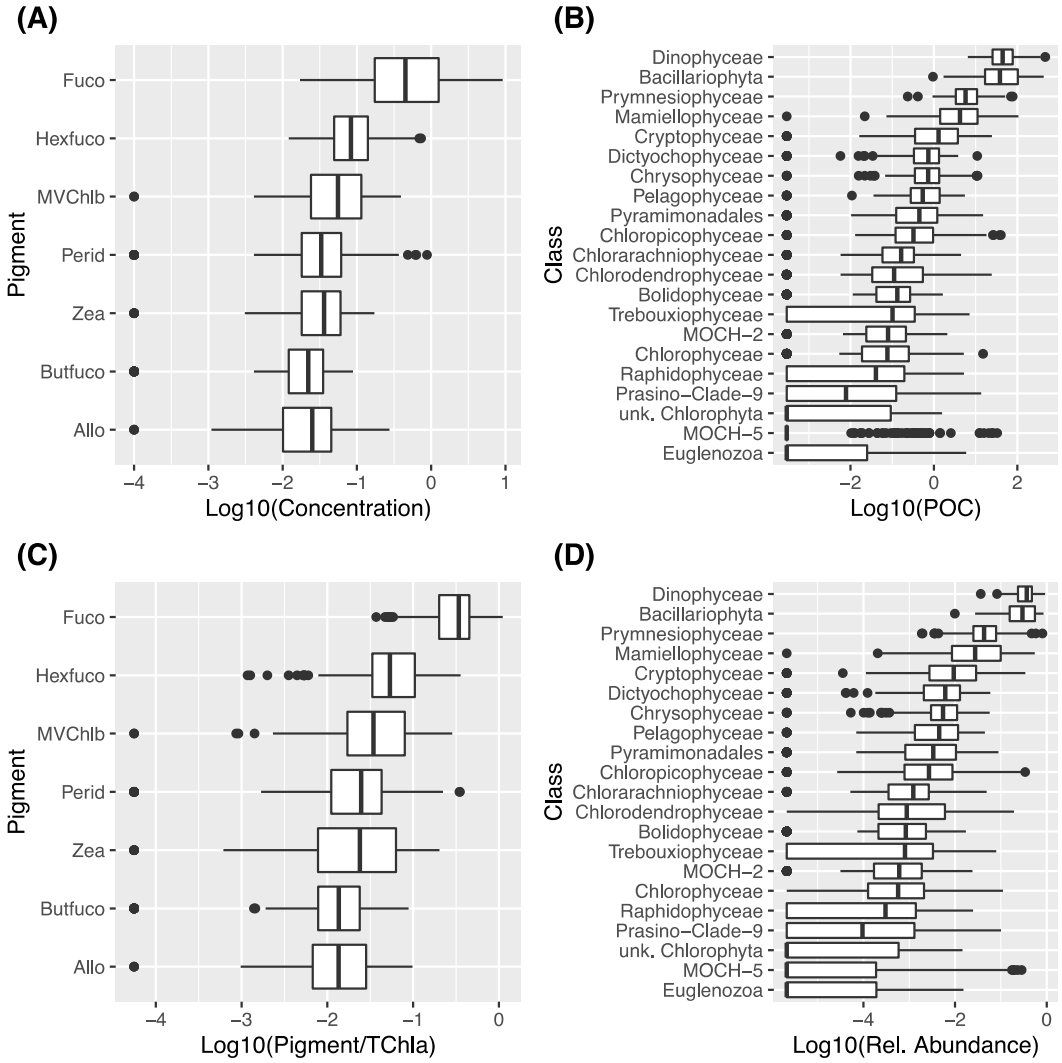

Figure 1. Distributions of (A) phytoplankton biomarker pigment concentrations, (B) phytoplankton class POC concentrations, (C) biomarker pigment ratios to total chlorophyll *a*, and (D) phytoplankton class relative sequence abundances observed by (A, C) HPLC pigment and (B, D) DNA meta-barcoding analysis. (B, D) only consider phytoplankton classes that comprise > 1% relative sequence abundance in at least one sample. Phytoplankton classes and biomarker pigments are sorted according to the rank order of their median relative sequence abundances and ratios to total chlorophyll *a*, respectively.

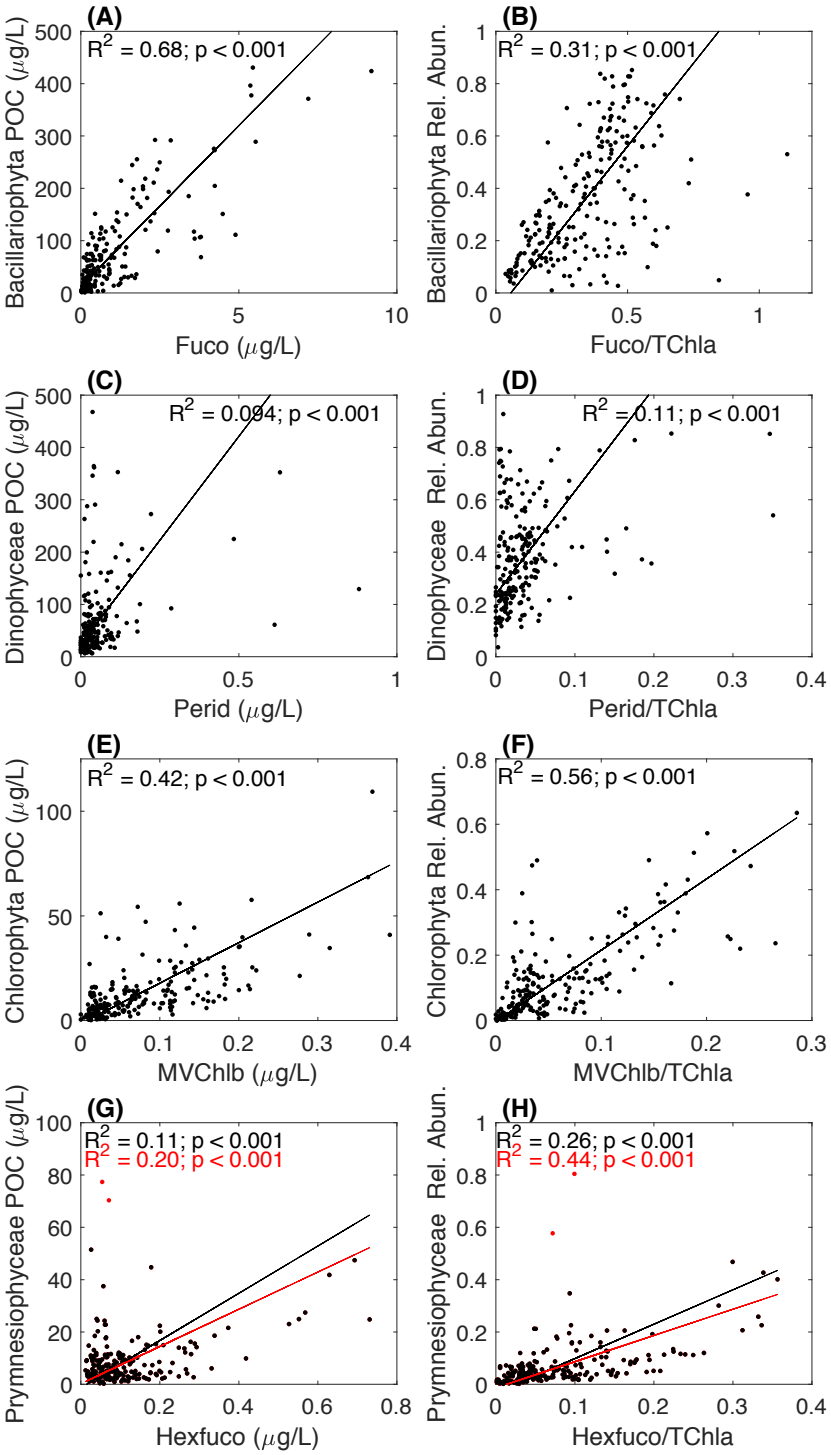

Figure 2. Comparisons of HPLC pigment and DNA meta-barcoding (A, C, E, G) biomass and (B, D, F, H) biomass proportion determinations for (A, B) diatoms, (C, D) dinoflagellates, (E, F) chlorophytes, and (G, H) prymnesiophytes. Values of squared Pearson correlation coefficients for each relationship are included in each panel. Lines of best fit are shown in black and were determined by type II linear regression analysis. In (G, H), red points show two outlier observations (see main text); these points were omitted for a second correlation analysis indicated with red text.

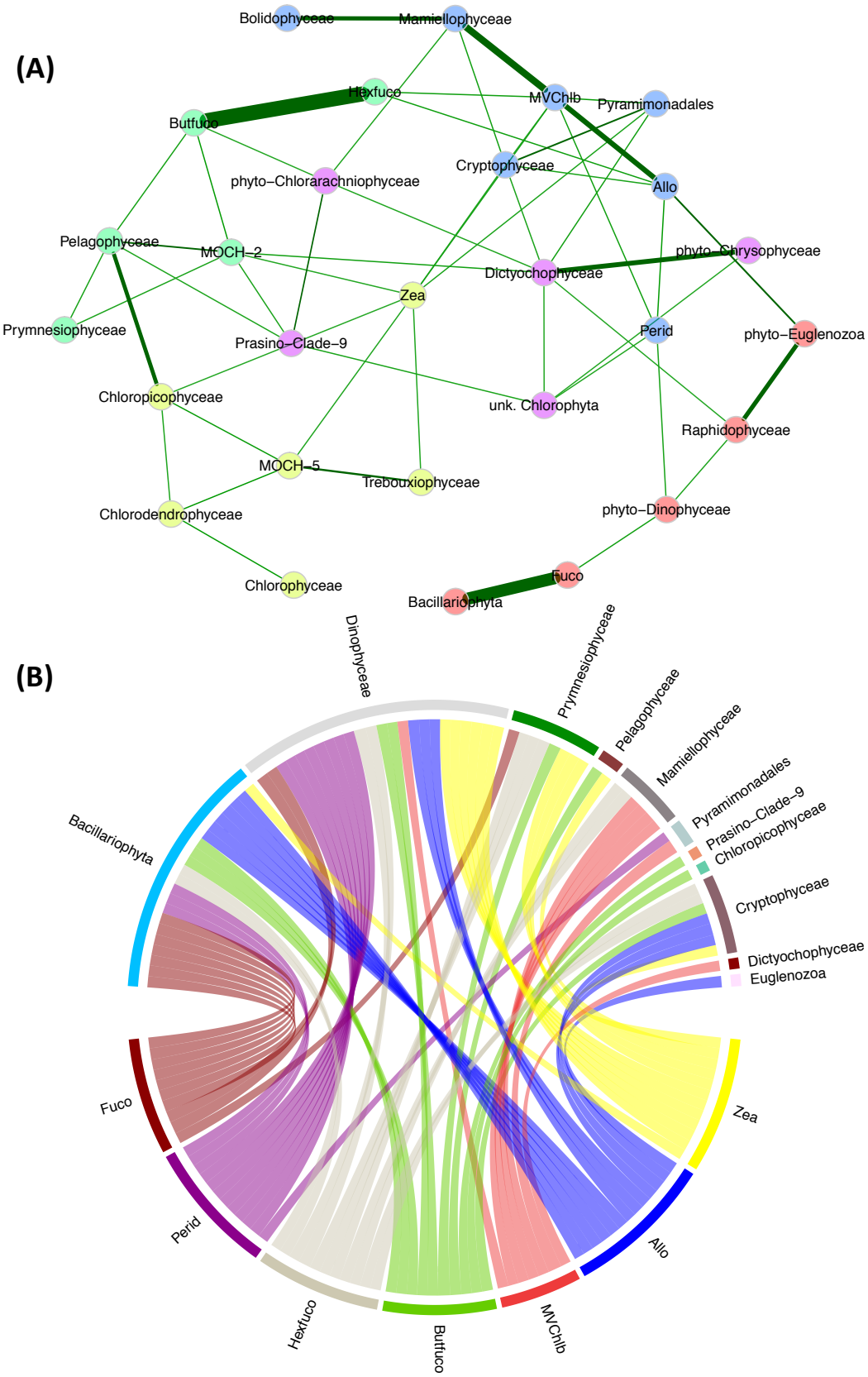
(A)
Bolidophyceae
Mamiellophyceae
Hexfuco
MVChlb
Pyramimonadales
Butfuco
Cryptophyceae
phyto−Chlorarachniophyceae
Allo
Pelagophyceae
MOCH−2
phyto−Chrysophyceae
Dictyochophyceae
Zea
Prymnesiophyceae
Perid
phyto−Euglenozoa
Prasino−Clade−9
Chloropicophyceae
unk. Chlorophyta
Raphidophyceae
MOCH−5
Trebouxiophyceae
Chlorodendrophyceae
phyto−Dinophyceae
Chlorophyceae
Fuco
Bacillariophyta
(B)
Dinophyceae
Prymnesiophyceae
Pelagophyceae
Mamiellophyceae
Pyramimonadales
Prasino−Clade−9
Chloropicophyceae
Cryptophyceae
Dictyochophyceae
Euglenozoa
Zea
Allo
MVChlb
Butfuco
Hexfuco
Perid
Fuco
Bacillariophyta

Figure 3. Associations amongst phytoplankton biomarker pigment concentrations and phytoplanktonic classes and ASVs determined from the covariation network including both phytoplankton and other protists. (A) shows the phytoplankton class POC and biomarker pigment concentration subnetwork. Nodes correspond to individual phytoplankton classes or biomarker pigment concentrations and are colored according to their community membership determined by an agglomerative community detection algorithm (Clauset et al., 2004). Edge thickness indicates the relative strength of association between nodes. All edges show positive associations. (B) shows positive associations between the seven diagnostic pigment concentrations and phytoplankton ASV POC concentrations.

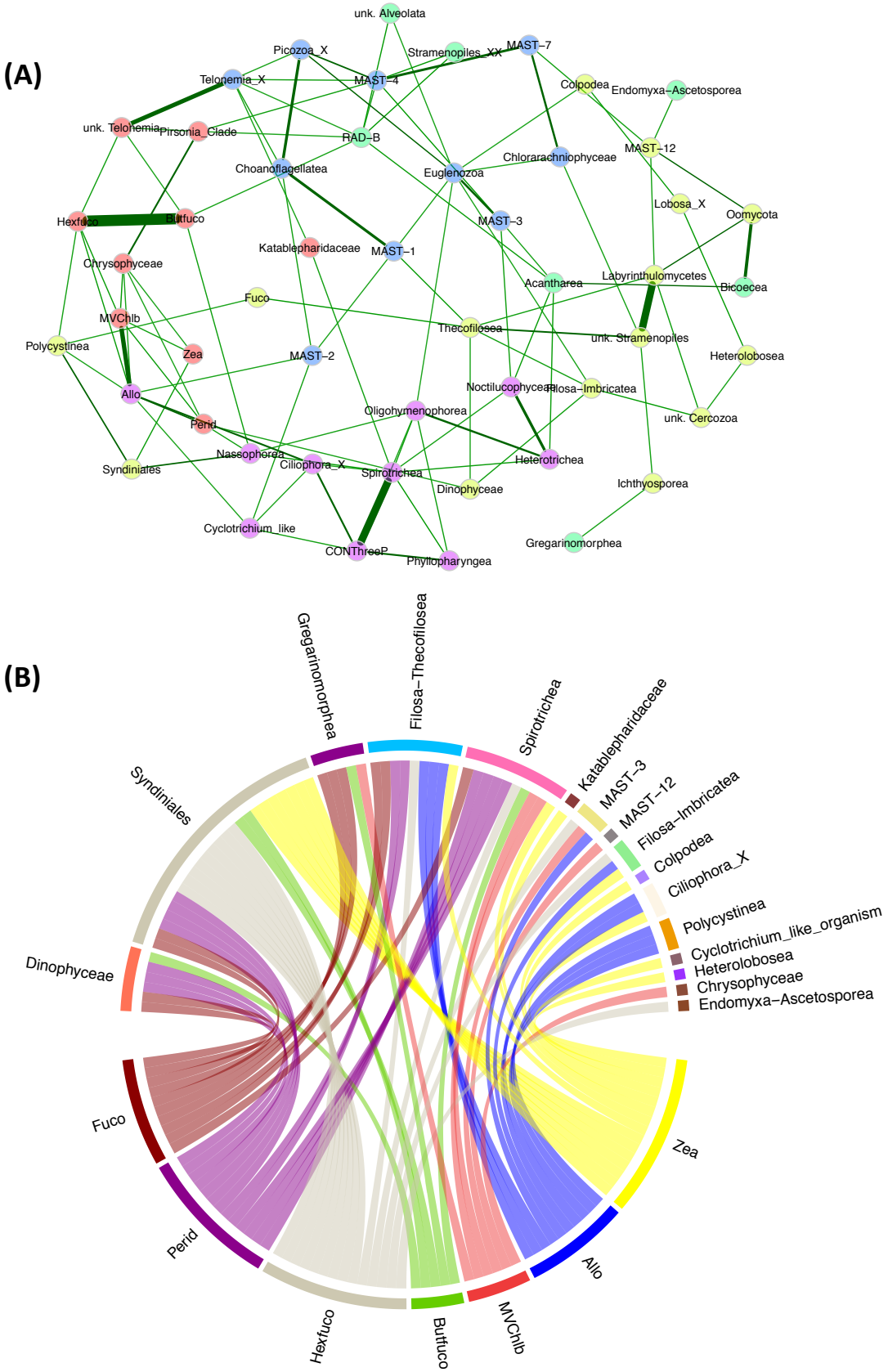
(A)
unk. Alveolata
Picozoa_X
Stramenopiles_XX
MAST−7
Telonemia_X
MAST−4
Colpodea
Endomyxa−Ascetosporea
unk. Telonemia
Pirsonia_Clade
RAD−B
MAST−12
Chlorarachniophyceae
Choanoflagellatea
Euglenozoa
Lobosa_X
Oomycota
Hexfuco
Butfuco
MAST−3
Katablepharidaceae
MAST−1
Chrysophyceae
Labyrinthulomycetes
Acantharea
Bicoecea
Fuco
MVChlb
Thecofilosea
unk. Stramenopiles
Polycystinea
Zea
MAST−2
Heterolobosea
Noctilucophyceae
Filosa−Imbricatea
Allo
Oligohymenophorea
unk. Cercozoa
Perid
Nasophorea
Heterotrichea
Syndiniales
Ciliophora_X
Spirotrichea
Dinophyceae
Ichthyosporea
Cyclotrichium_like
CONThreeP
Gregarinomorphea
Phyllopharyngea
(B)
Gregarinomorphea
Filosa−Thecofilosea
Spirotrichea
Katablepharidaceae
MAST−3
MAST−12
Filosa−Imbricatea
Colpodea
Ciliophora_X
Polycystinea
Cyclotrichium_like_organism
Heterolobosea
Chrysophyceae
Endomyxa−Ascetosporea
Syndiniales
Dinophyceae
Fuco
Perid
Hexfuco
Butfuco
MVChlb
Allo
Zea

Figure 4. Associations amongst phytoplankton biomarker pigment concentrations and non-phytoplanktonic classes and ASVs determined from the covariation network including both phytoplankton and other protists. As in Figure 3, (A) shows the class POC and biomarker pigment concentration covariation subnetwork. Nodes correspond to individual classes or biomarker pigment concentrations and are colored according to their community membership determined by an agglomerative community detection algorithm (Clauset et al., 2004). Edge thickness indicates the relative strength of association between nodes. All edges represent positive associations. (B) shows positive associations between the seven diagnostic pigment concentrations and non-phytoplanktonic ASV POC concentrations.

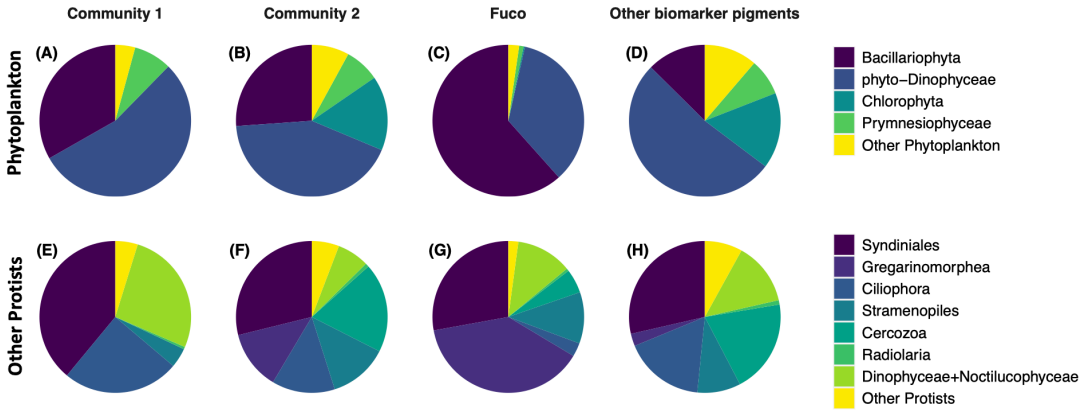

Figure 5. Mean relative sequence abundances of (A-D) phytoplankton and (E-H) other protistan ASVs found within the four communities identified by the community detection algorithm (Clauset et al., 2004). Communities 1 and 2 do not include any of the seven diagnostic pigments considered in network analysis. "Other Phytoplankton" includes ASVs representative of *Euglenozoa*, *Chrysophyceae*, *Dictyophyceae*, *Cryptophyceae*, *Raphidophyceae*, *Chlorarachniophyceae*, *Pelagophyceae*, and MOCH, while "Other Protists" includes *Lobosa*, *Discoba*, *Katablepharidophyta*, *Picozoa*, *Mesomycetozoa*, *Telonemia*, *Chrysophyceae*, unknown *Alveolata*, and *Choanoflagellatea*.

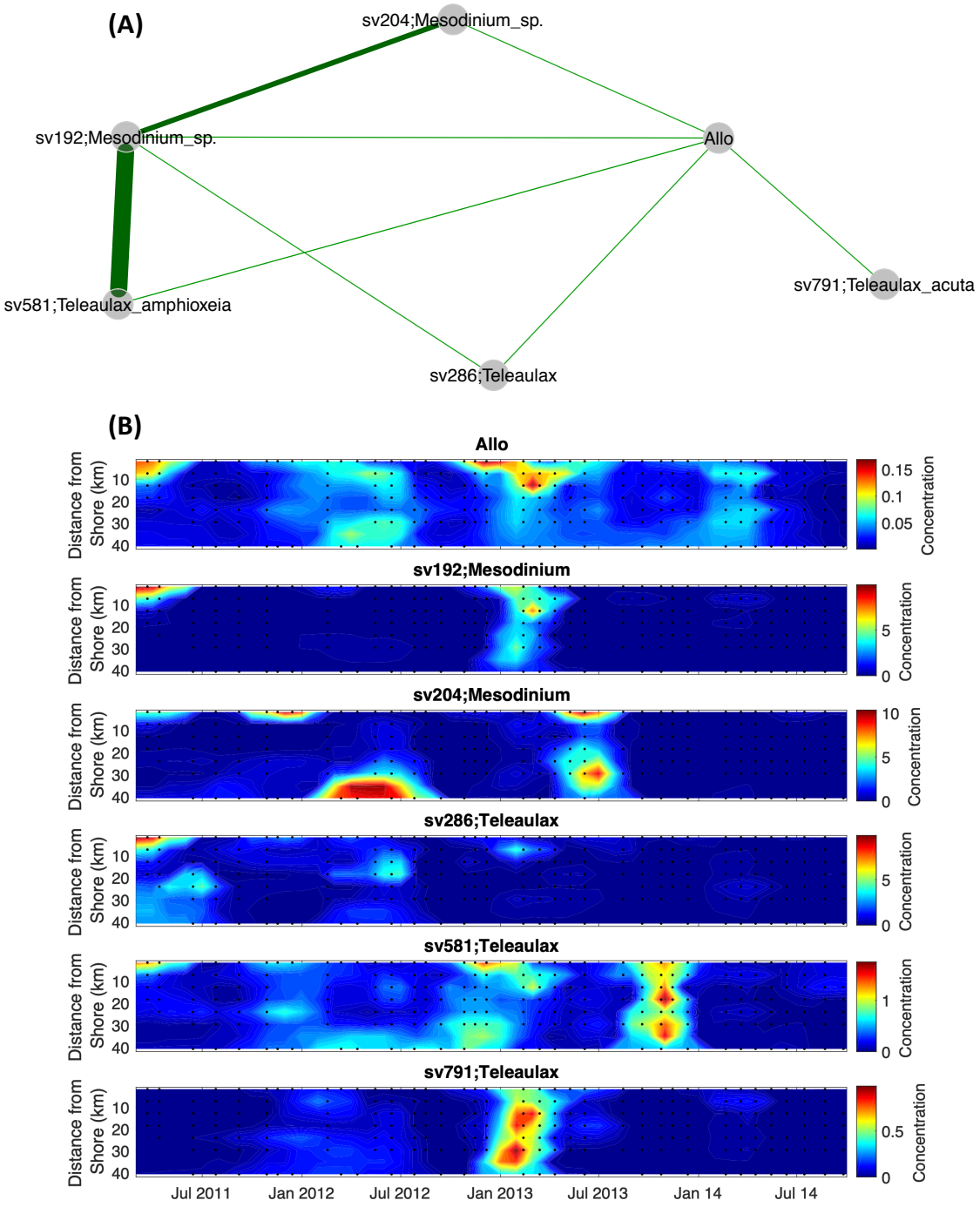

Figure 6. Example of a novel perspective of phytoplankton and protistan ecology provided by integrating HPLC pigment and DNA meta-barcoding analyses. (A) shows a network motif including the cryptophyte biomarker pigment Allo and several associated protistan ASVs, including putative *Mesodinium* ciliates and their putative cryptophyte prey/symbiont, *Teleaulax*, from which they can acquire the capacity for photoautotrophy. (B) shows the spatiotemporal dynamics of Allo in addition to each of the ASVs included in the network motif in (A).

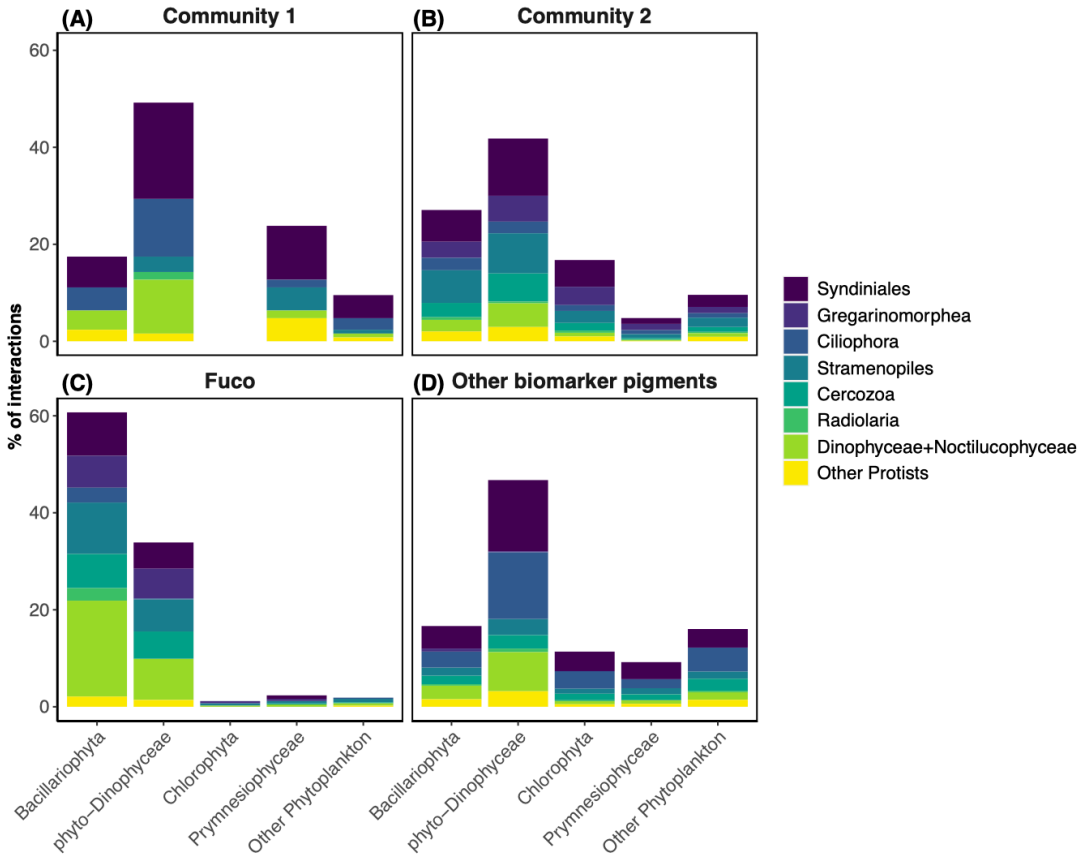

Figure 7. Composition of interactions amongst each of the four communities identified by the community detection algorithm (Clauset et al., 2004) applied to the protistan ASV and biomarker pigment concentration network. "Other Phytoplankton" includes ASVs representative of *Euglenozoa*, *Chrysophyceae*, *Dictyophyceae*, *Cryptophyceae*, *Raphidophyceae*, *Chlorarachniophyceae*, *Pelagophyceae*, and MOCH, while "Other Protists" includes *Lobosa*, *Discoba*, *Katablepharidophyta*, *Picozoa*, *Mesomycetozoa*, *Telonemia*, *Chrysophyceae*, unknown *Alveolata*, and *Choanoflagellatea*.

V. Appendices

Appendix 1. Supplemental Information for Chapter II

Introduction

The Supplemental Information contains additional text to clarify the procedures used to select a smoothing filter for use in the derivative analysis of phytoplankton absorption coefficients (Text S1), in the CHEMTAX analysis of phytoplankton pigment concentrations (Text S2), in the derivation of the coefficients used in the modeling exercise (Text S3), and in the derivation of the indices of community composition based on CHEMTAX and diagnostic pigment (DP) analyses that we attempted to model (Text S4). We include results of the smoothing filter optimization procedure employed for the derivative analysis of the phytoplankton absorption coefficient (Table S1, Figure S1), results of preliminary cluster analyses (Figure S2), and the amplitude functions (AFs) of the first four pigment empirical orthogonal function (EOF) modes and their correlations with several environmental parameters to aid in their ecological interpretation (Table S2, Figure S3; see Figure 2 in main text for loadings) of the phytoplankton pigment data set to support our interpretations of these results in the main text. Finally, we include the results (Figure S4) and final optimized pigment ratio matrices (Tables S3-11) from the CHEMTAX calculations to demonstrate that our results have been quality controlled according to the guidelines presented in Higgins et al. (2011), the multiple linear regression (MLR) coefficients used in the DP analyses based on the approach of Uitz et al. (2006) (Table S12), the results of an evaluation of the sensitivity of the modeling approach to assumptions used in its formulation (Table S13; see Text S3 for more information), all goodness of fit statistics for 500-permutation cross-validations of all pigments and indices of community composition we attempted to model with the phytoplankton absorption coefficient derivative spectra (Table S14), and all goodness of fit statistics for 500-permutation model cross-validations of selected pigment indices when the modeling approach is applied to the total non-water absorption coefficient (Table S15).

Text S1. Derivative analysis smoothing filter optimization

We investigated four different smoothing filter types including the moving average, Savitsky-Golay, Lowess, and Hamming window filters. The moving average and Savitsky-Golay filters have been described in detail previously (Tsai and Philpot, 1998; Vaiphasa, 2006). The Lowess filter uses weighted linear least squares to fit a first order polynomial to consecutive subsets of the data as identified by the length of the filter (Cleveland, 1981). The Hamming window filter, which is used as a form of Fourier filtering, was applied by convolving each spectrum with the Hamming window function represented by:

$$w(n) = 0.54 - 0.46 * \cos\left(\frac{2*\pi*n}{N}\right) \quad (1)$$

where N is the length of the filter and n corresponds to the n^{th} point of the filter (i.e. $0 \leq n \leq N - 1$; Harris, 1978).

Past work shows that there is variability in the positions of specific pigment absorption maxima between studies, even when similar analytical methods are employed (Bidigare et al., 1989; Bricaud et al., 2004; Chase et al., 2013; Lohrenz et al., 2003). Additionally, some pigments have absorption maxima very close to one another meaning that instrument noise and the application of smoothing filters makes it very difficult to resolve these peaks independently from one another using a derivative analysis. Therefore, for each absorption signature considered, a spectral filter width optimization procedure was constructed. First, the minimum in the second derivative spectra over a pre-determined wavelength range containing only the absorption peak or peaks of interest was identified. Then, the pigment or sum of all pigments with a maximum in absorption within this wavelength range was regressed against the value of the second derivative at this position. This procedure was repeated for six pairs of pigments and absorption maxima (Table S1) to identify candidate smoothing filter types and sizes with a focus on selecting the smallest filter size that yielded strong linear relationships with pigments (e.g., Figure S1). We then applied each candidate filter to independent derivative analyses, examined the strength of linear relationships of all pigments with the derivative spectra at each wavelength (see Figure 3 in main text), and qualitatively evaluated the resulting derivative spectra.

Text S2. CHEMTAX analysis details

The CHEMTAX program is a widely-used method for assessing the contributions of phytoplankton taxa to TChla (Higgins et al., 2011; Mackey et al., 1996). CHEMTAX decomposes a suite of pigment concentrations into a matrix of the relative contributions of a pre-selected set of phytoplankton taxa to TChla by optimizing an initial guess of the taxon-specific pigment ratios (Mackey et al., 1996). Because pigment ratios can vary dramatically, it is necessary to partition large data sets into subsets assumed to have similar pigment ratios to reduce uncertainties introduced by the CHEMTAX calculation (Higgins et al., 2011 and references therein; Swan et al., 2016).

We performed separate CHEMTAX analyses on the 10% most extreme positive and negative AFs of each of the first four pigment EOF modes, and an analysis on all remaining data not found in the 10% negative or positive bounds of any mode, to help reduce uncertainty due to changing pigment ratios throughout the time series (similar to Anderson et al., 2008). If separate fractional contributions were computed for the same data point (i.e. a data point was found in more than one data subset), the outputs were averaged. All calculations used the pigment ratio optimization procedure outlined in Swan et al. (2016). The pigment groups, or taxa, used in all initial CHEMTAX calculations were DIATOM-1, HAPTO-6, HAPTO-8, DINO-1, CRYPTO-1, PELAGO-1, CHLORO-1, and PRASINO-1 (Higgins et al., 2011). All pigments used in the EOF analysis except for DVChla and Pras were initially included in all analyses. We included Pras and PRASINO-3, and CYANO-4, when Pras and DVChla, respectively, appeared to play a significant role in the data subset of interest. Initial pigment ratio matrices were constructed using mean literature values from field data according to Higgins et al. (2011). If field data were unavailable or sparse, laboratory data was used. Pigments with mean RMSE's larger than 50% and phytoplankton classes with final pigment ratios

outside of realistic bounds in the initial analyses were discarded and the calculation was repeated until all pigments included had mean RMSE's less than 50% and the final ratios prescribed to each phytoplankton class were within the bounds presented in Higgins et al. (2011). Tables detailing the pigments, phytoplankton classes, and final optimized pigment ratios used in the CHEMTAX calculations for each data subset are presented in Tables S3-S11, and the results of these analyses are presented in Figure S4.

The CHEMTAX results (Figure S4) were in rough agreement with the results of the EOF analysis. For each EOF mode-specific subset of pigment data, the taxon with the largest average contribution to TChla derived from CHEMTAX agreed with the taxonomic interpretation of the EOF mode. However, within each data subset there was often a large range of fractional contributions to TChla meaning a taxon thought to be insignificant for a given EOF mode dominated at a few extreme occurrences of that mode. Direct comparisons of these results are complicated since the EOF modes frequently capture the covariance of two or more taxa (e.g., mode 2 at positive AFs corresponds to mixed nanoplankton, picoplankton, and haptophytes), while CHEMTAX assumes that abundances of specific taxonomic groups are uncorrelated (Higgins et al., 2011). Phytoplankton pigments are clearly not independent from each other based upon the results of the EOF and cluster analyses above, which undermines the validity of CHEMTAX analysis when applied to this data set.

Text S3. Details on the formulation and validation of the linear model

Principal components regression (PC-R) is sensitive to the number of components used as inputs to the multiple linear regression. Bracher et al. (2015), who used a PC-R-based approach with remote sensing reflectance observations, started with a model using a large number of principal components and used a stepwise regression procedure based on minimizing the Aikake Information Criterion (AIC) to reduce the number of components used in their model. We found that automated stepwise regression procedures using the AIC or other criteria to select components for use in the multiple linear regression did not perform well, possibly because these approaches often resulted in removal of components that explained a relatively large proportion of the variability in $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$. Therefore we assumed that principal components explaining a larger proportion of the variability than the component that explained the minimum variability, but was included in the model, were essential, and that the first 100 principal components of $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ contained all of the information relevant to modeling all pigments.

To test the sensitivity of the model's performance to the assumption that the first 100 principal components of $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ contained all of the information relevant to modeling all pigments, we cross-validated the model while allowing a maximum of 25, 50, 100, 150, and 200 principal components to be used in it's formulation. This sensitivity test was conducted using two unique biomarker pigments: one which was retrieved extremely well (Fuco; $R^2 > 0.8$ for all cross-validations), and one which was retrieved with intermediate levels of success (HexFuco; $0.75 > R^2 > 0.64$ across all cross-validations). The results from these sensitivity tests are shown in Table S13. Marginal improvements in R^2 were observed as the maximum number of components allowed in the model increased. However, percent error statistics showed variable responses to these

changes, and the model coefficients became obscured by noise so that very few, if any, bands were significant (e.g., Figure 6) when the maximum number of components allowed in the model was too large, likely because higher order principal components are generally dominated by noise. We found that allowing a maximum of 100 principal components in the model offered the best balance in limiting the noise in the model coefficients while accurately and precisely retrieving each p_m.

Under these assumptions, we used a 5-fold cross-validation procedure to derive each set of $A_m(\lambda)$ and $B_m(\lambda)$. The cross-validation procedure consisted of: 1. Randomly dividing the training data set into five subsets; 2. Training a model using the first principal component of four of the training data subsets; 3. Validating the model from step 2 using the fifth subset; 4. Sequentially adding principal components and repeating steps 2 and 3 for models containing 1 to 100 principal components; 5. Selecting the best model of those evaluated in step 4 as determined by maximizing R^2; and 6. Repeating steps 2-5 five times so that each data subset was used as the validation subset once. This procedure resulted in five subsets of $A_m(\lambda)$, $B_m(\lambda)$, and C_m, which were averaged to derive the final coefficients of the model. To assess the performance of our modeling approach, we used a validation procedure similar to that of Bracher et al., (2015) where the data were randomly divided into a training (75%) and a validation (25%) set 500 times and a new model was produced and validated each time.

It should be noted that all goodness of fit statistics for validating modeled fractional contributions to TChla were computed after constraining modeled observations to lie between 0 and 1, while modeled pigment concentrations were only constrained to be larger than 0 and modeled pigment mode amplitude functions were not constrained. Prior to computing percent error statistics for validation of the model, we added 0.0001 to all observed pigment concentrations of 0 to prevent undefined solutions.

Text S4. Derivation of modeled indices of community composition via CHEMTAX and Diagnostic Pigment (DP) analyses

To facilitate comparisons with pigment EOF modes, relative contributions to TChla derived from CHEMTAX were summed to broader taxonomic groupings as follows: diatoms = DIATOM-1; haptophytes = HAPTO-6 + HAPTO-8; dinoflagellates = DINO-1; nanoplankton = CRYPTO-1 + PELAGO-1 + CHLORO-1 + PRASINO-1; picoplankton = PRASINO-3 + CYANO-4. We also conducted several DP analyses following the approaches of Vidussi et al. (2001) and Uitz et al. (2006), and attempted to model the fractional contributions of each taxon and size class to TChla as derived from these analyses. The Vidussi et al. (2001) approach uses unweighted ratios of seven DP's to derive the "biomass proportion" of each taxon assumed to be represented by each of the seven DP's. The biomass proportion of three size classes, defined as sums of the biomass proportions of individual DP's, can also be derived with this method. The approach of Uitz et al. (2006) weighted each of the seven DP's defined by Vidussi et al. (2001) based on their regression coefficients in a multiple linear regression (MLR) of the DP's against TChla. Uitz et al. (2006) used this approach on a global data set; we therefore evaluated our model's ability to retrieve the DP contributions to TChla derived via the global MLR coefficients presented by Uitz et al. (2006), as well as a "locally tuned" set of

coefficients. To derive the locally tuned set of coefficients, we carried out an MLR of the seven DP's against TChla and found that the coefficient of Zea was less than 10^{-4} for this dataset. We therefore excluded Zea in a second MLR analysis, and used the fractional contributions derived using the second MLR analysis to derive "locally-tuned" fractional contributions to TChla. The coefficients used in both MLR analyses are listed in Table S12.

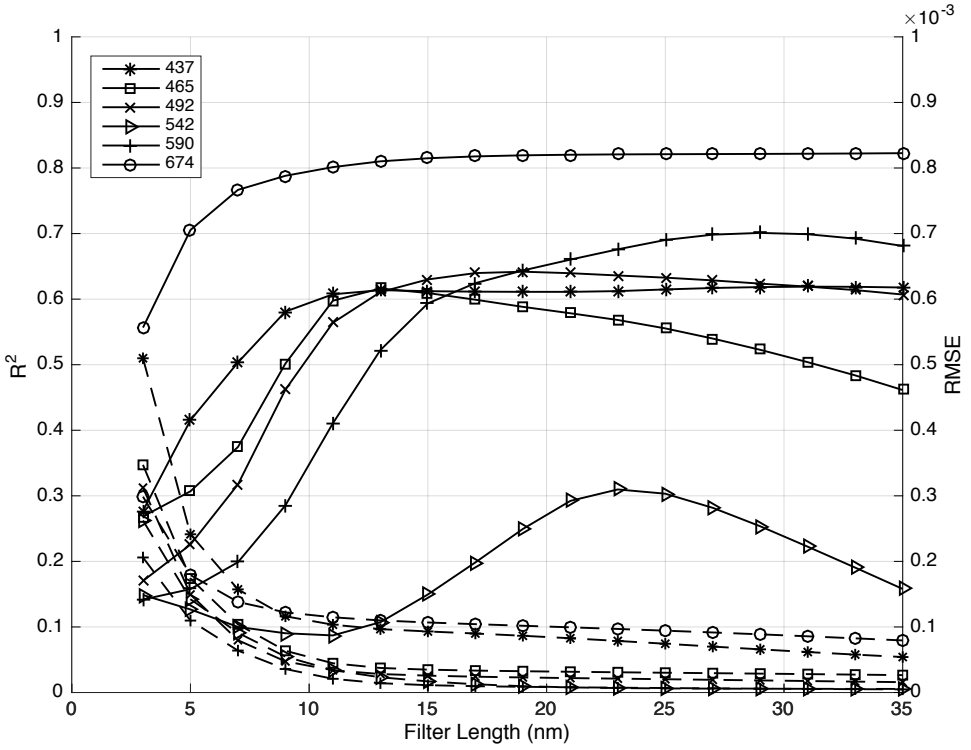

Figure S1. Example of derivative analysis smoothing filter optimization procedure. Statistics for the Hamming window filter are shown, with solid lines showing R^2 and dashed lines showing RMSE. The numbers in the legend correspond to the wavelength of maximum absorption on average of the pigment or pigment group specified in Table S1.

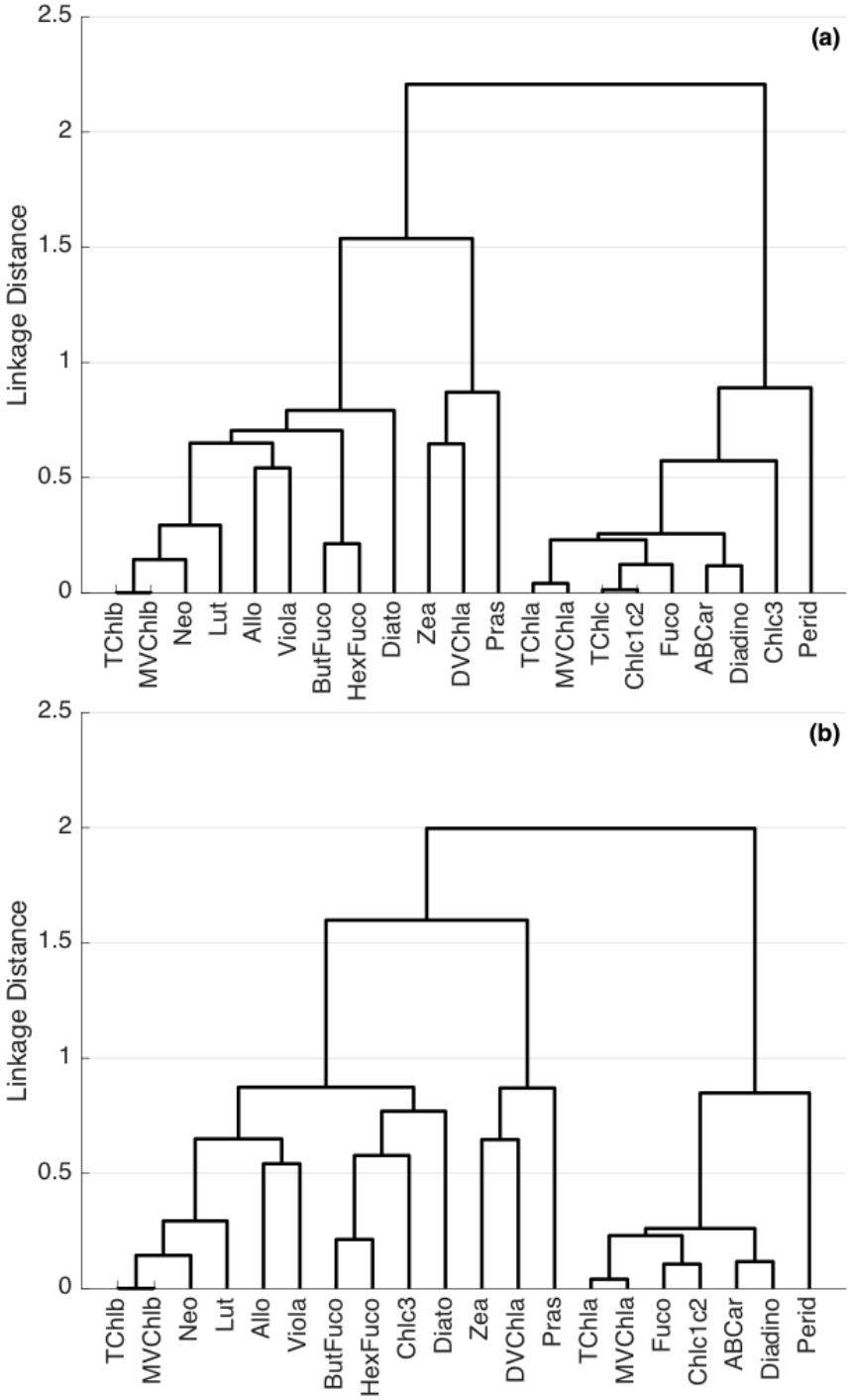

Figure S2. Results of the two preliminary analyses using the correlation distance and Ward's linkage method. (a) includes TChlc, MVChlb, and MVChla, (b) includes MVChlb and MVChla, and the final analysis (Figure 1 in main text) excludes all three of these pigments. Note the change in position of Chlc3 when TChlc is removed from the analysis.

Figure S3. Amplitude functions corresponding to the first four phytoplankton pigment EOF modes (see Figure 2 in main text for EOF loadings).

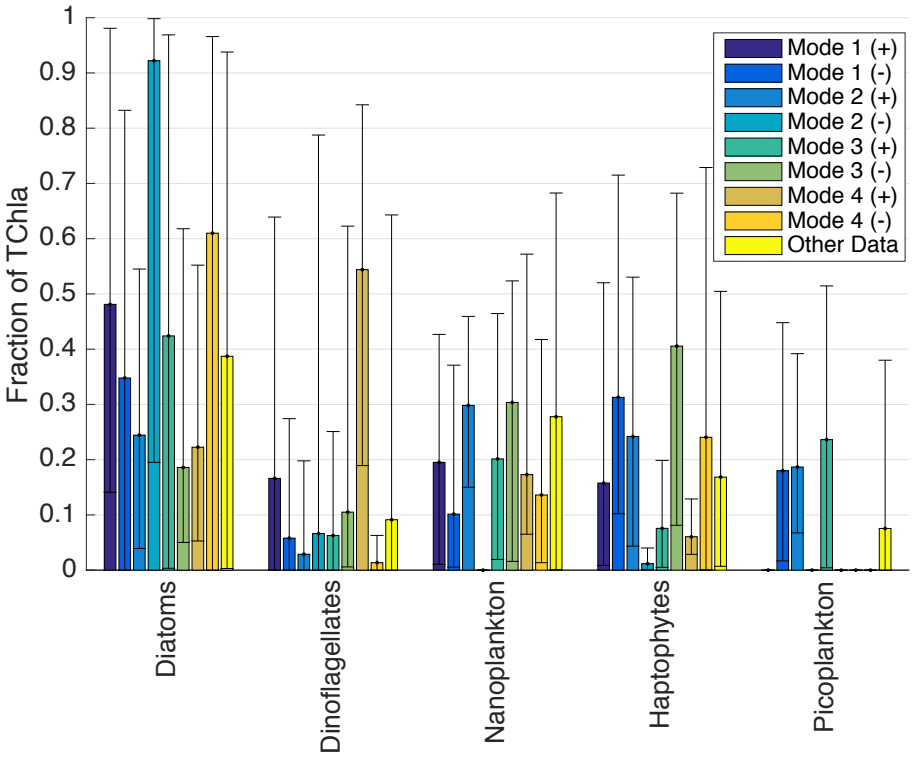

Figure S4. Fractional contributions of each major taxonomic group to TChla derived from CHEMTAX for each data subset partitioned by pigment EOF mode (see Text S2 for details on data partitioning). Bars show average contributions and error bars indicate the minimum and maximum contributions in each data subset. Contributions of zero result from no representatives of that taxon being included in the CHEMTAX calculation due to their removal according to the quality control procedure (described in Text S2).

Target Peak	Wavelength Range in Literature[a]	Allowed Wavelength Range	Contributing Pigments
440	435-440	430-448	TChla
465	460-470	455-475	TChlb + TChlc + Zea + Diadino + Diato + Allo + ABCar
490	490-492	480-500	Zea + Diadino + Diato + Allo + ABCar[b]
530	521-539	515-545	Fuco + Perid[b]
585	585-586	580-600	TChlc
675	675	670-680	TChla

Table S1. Pigments and corresponding absorption signatures used in the smoothing filter optimization procedure. All wavelengths are in nm.
[a]References: Bidigare et al. (1990), Lohrenz et al. (2003), Bricaud et al. (2004), Chase et al. (2013).
[b]Some literature suggests the photosynthetic carotenoids (Fuco, Perid, HexFuco, ButFuco) should all contribute to absorption features near 490 and 530 nm. However, we found that including only the listed pigments in the regression produced stronger linear relationships.

Table S2. Taxonomic interpretations of the first four pigment EOF modes and the correlation coefficients of each mode with environmental parameters. Insignificant correlations at 95% confidence are noted as N.S. and the number of observations used in each calculation is noted next to each environmental parameter.

Table S3. Final optimized pigment ratios for the CHEMTAX calculation on the data subset corresponding to the 10% largest positive amplitude functions of pigment EOF mode 1.

Table S4. Final optimized pigment ratios for the CHEMTAX calculation on the data subset corresponding to the 10% largest negative amplitude functions of pigment EOF mode 1.

Table S5. Final optimized pigment ratios for the CHEMTAX calculation on the data subset corresponding to the 10% largest positive amplitude functions of pigment EOF mode 2.

Table S6. Final optimized pigment ratios for the CHEMTAX calculation on the data subset corresponding to the 10% largest negative amplitude functions of pigment EOF mode 2.

Table S7. Final optimized pigment ratios for the CHEMTAX calculation on the data subset corresponding to the 10% largest positive amplitude functions of pigment EOF mode 3.

Table S8. Final optimized pigment ratios for the CHEMTAX calculation on the data subset corresponding to the 10% largest negative amplitude functions of pigment EOF mode 3.

Table S9. Final optimized pigment ratios for the CHEMTAX calculation on the data subset corresponding to the 10% largest positive amplitude functions of pigment EOF mode 4.

Table S10. Final optimized pigment ratios for the CHEMTAX calculation on the data subset corresponding to the 10% largest negative amplitude functions of pigment EOF mode 4.

Table S11. Final optimized pigment ratios for the CHEMTAX calculation on the remaining data that was not found in the 10% largest positive or negative amplitude functions of the first four pigment EOF modes.

Biomarker Pigment	Global	Local, with Zea	Local, without Zea
Fuco	1.41	1.688	1.688
Perid	1.41	1.740	1.741
Allo	0.60	2.663	2.659
ButFuco	0.35	2.885	3.058
HexFuco	1.27	0.722	0.694
TChlb	1.01	3.185	3.172
Zea	0.86	4.49E-05	NA

Table S12. Coefficients from the multiple linear regression of the seven diagnostic pigments on TChla following Uitz et al. (2006). Global coefficients are taken directly from Uitz et al. (2006). "Local, without Zea" was used to derive the indices of community composition we attempted to model as the "locally tuned" indices of community composition.

Table S13. Results demonstrating the sensitivity of the model to the assumption that the first 100 principal components of the first and second derivative spectra contain all information relevant for modeling all phytoplankton pigment concentrations and indices of community composition.

Table S14. All goodness of fit statistics for the 500-fold cross-validations of all pigments and pigment-derived indices of community composition that we attempted to model using the phytoplankton absorption coefficient.

Table S15. All goodness of fit statistics for the 500-fold cross-validations of the pigments and pigment EOF modes that we attempted to model using the total non-water absorption coefficient.

Appendix 2. Supplemental Information for Chapter III

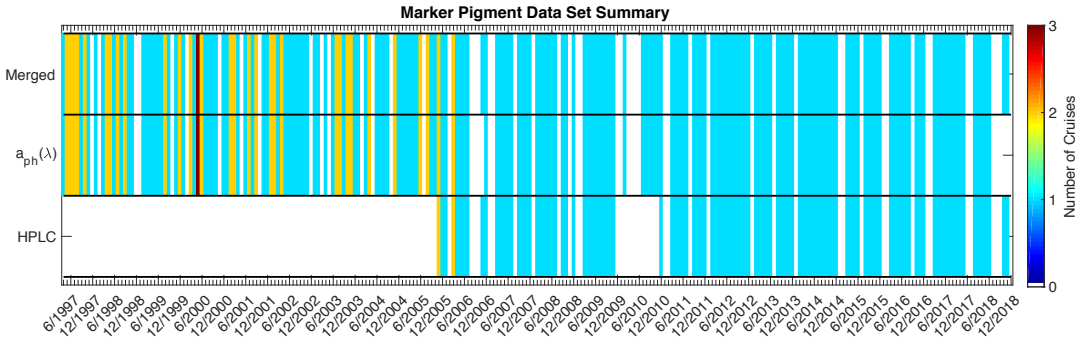

Figure S1. Merged biomarker pigment data set coverage. White tiles indicate no available observations in a particular month of the Plumes and Blooms time series for a particular data product (High Performance Liquid Chromatography pigment concentrations, HPLC; spectrophotometric phytoplankton absorption coefficient determinations, $a_{ph}(\lambda)$; the sum of independent HPLC and $a_{ph}(\lambda)$ observations, Merged).

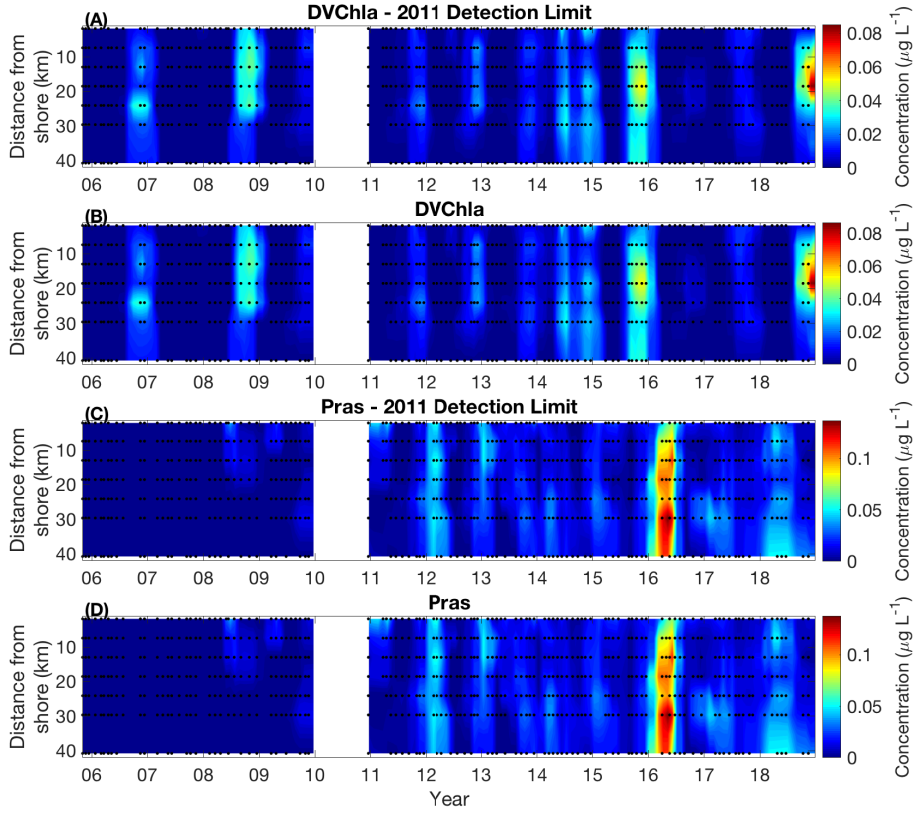

Figure S2. Time series of (A, B) DVChla and (C, D) Pras, (A, C) with and (B, D) without detection limit corrections to account for the change in detector settings implemented in 2014 that impacted all subsequent samples analyzed. Note that the laboratory where samples were analyzed changed for all samples collected from 2011 to the end of the record, but this change is not accounted for in these data.

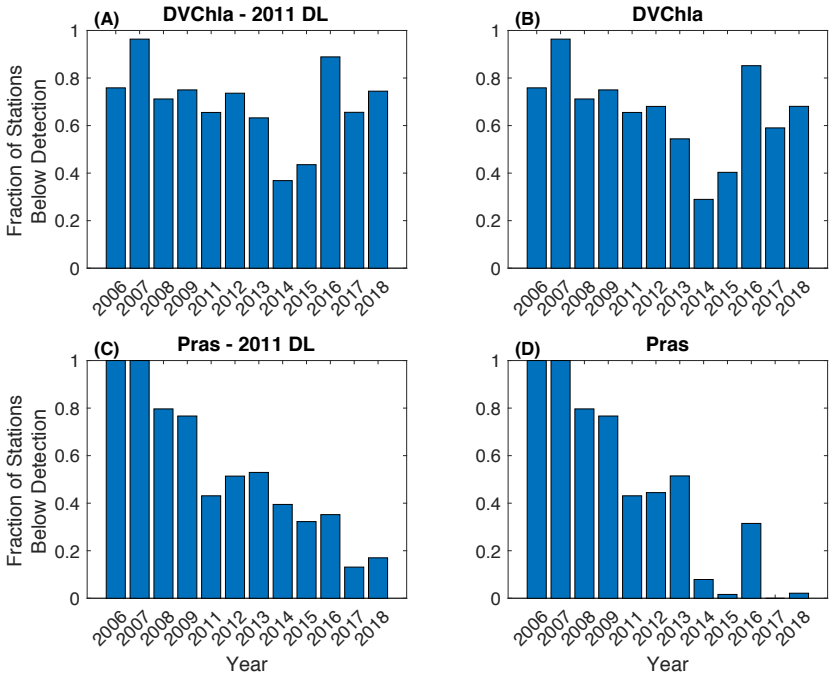

Figure S3. The fraction of total Plumes and Blooms stations with HPLC pigment samples available where (A, B) DVChla and (C, D) Pras were not detected (A, C) with and (B, D) without detection limit corrections to account for the change in detector settings implemented in 2014 that impacted all subsequent samples analyzed. 2005 and 2010 were omitted due to a small number of samples available in those years.

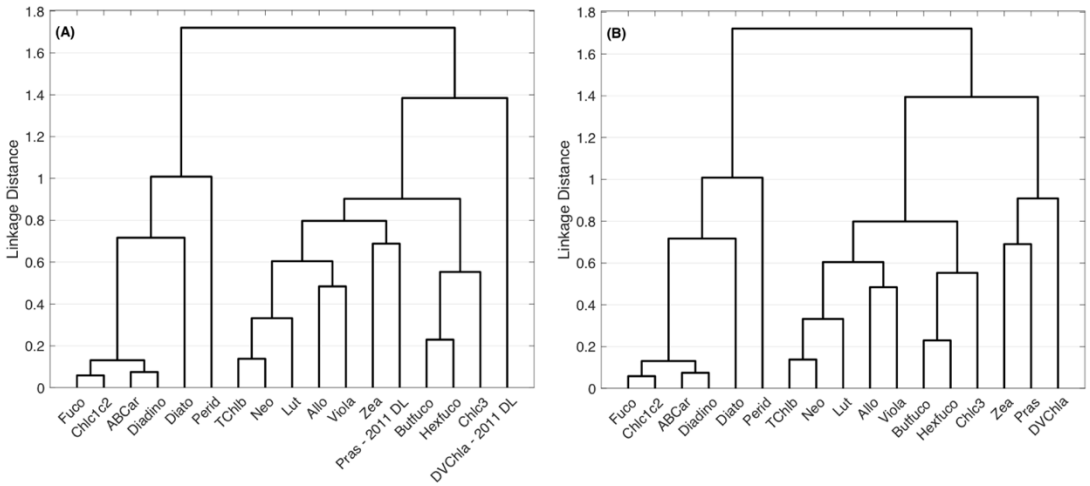

Figure S4. Hierarchical cluster analysis using correlation distance and Ward's linkage method of HPLC pigment data (A) with and (B) without detection limit corrections implemented on DVChla and Pras. As noted in the main text, the variable positioning of Pras led to its exclusion from consideration of the cluster analysis presented in the main text (Figure 2).

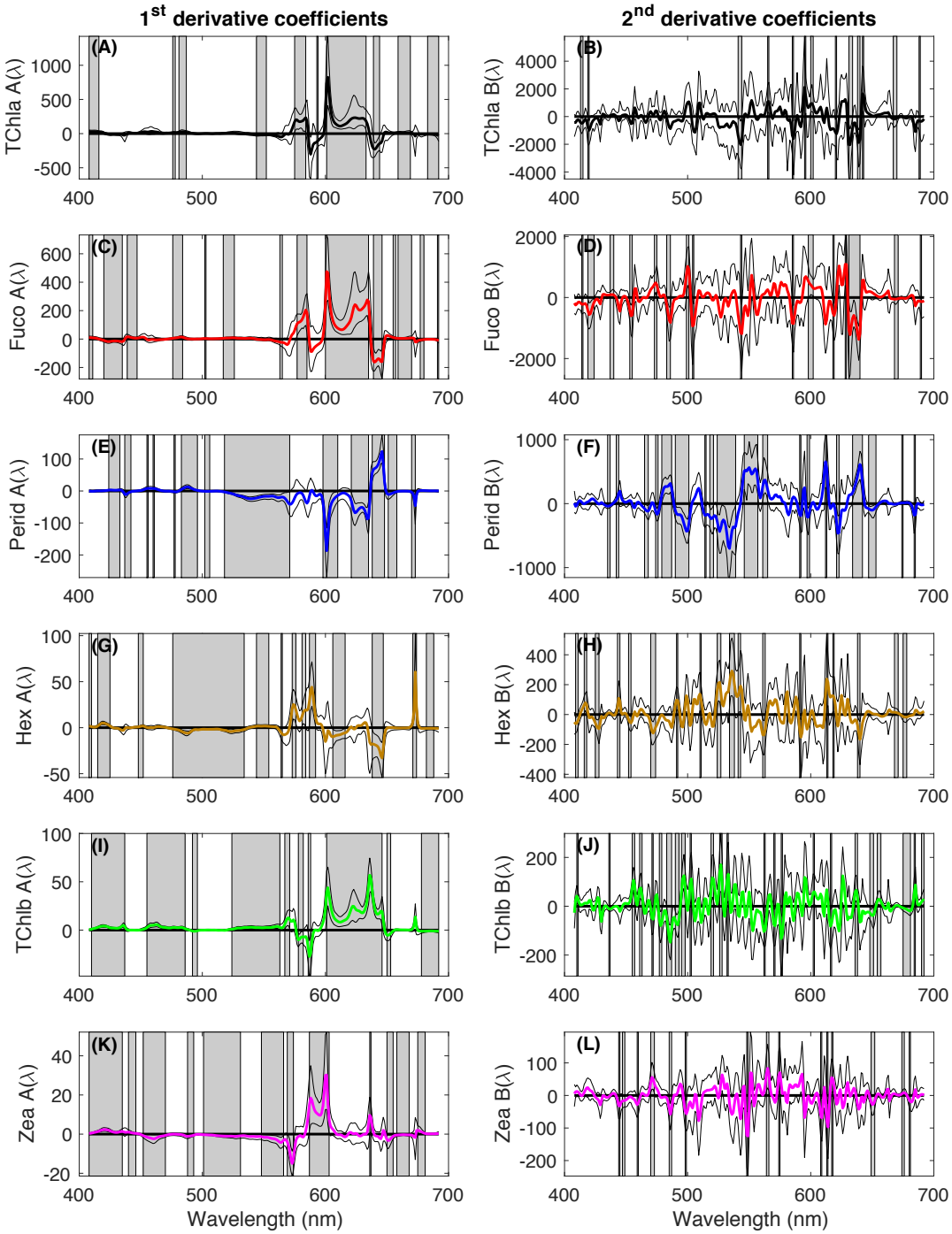

Figure S5. Mean (bold line color-coded to match the convention used in the main text) +/- empirical 95% confidence intervals (thin black lines) of optimized model coefficients for (A, C, E, G, I, K) first and (B, D, F, H, J, L) second spectral derivatives of phytoplankton absorption coefficients as determined in the 500-fold cross-validation exercise. Bands significantly different from zero with width >1 nm are shaded. The mean coefficients shown here were used to model each pigment according to equation 1 in the main text.

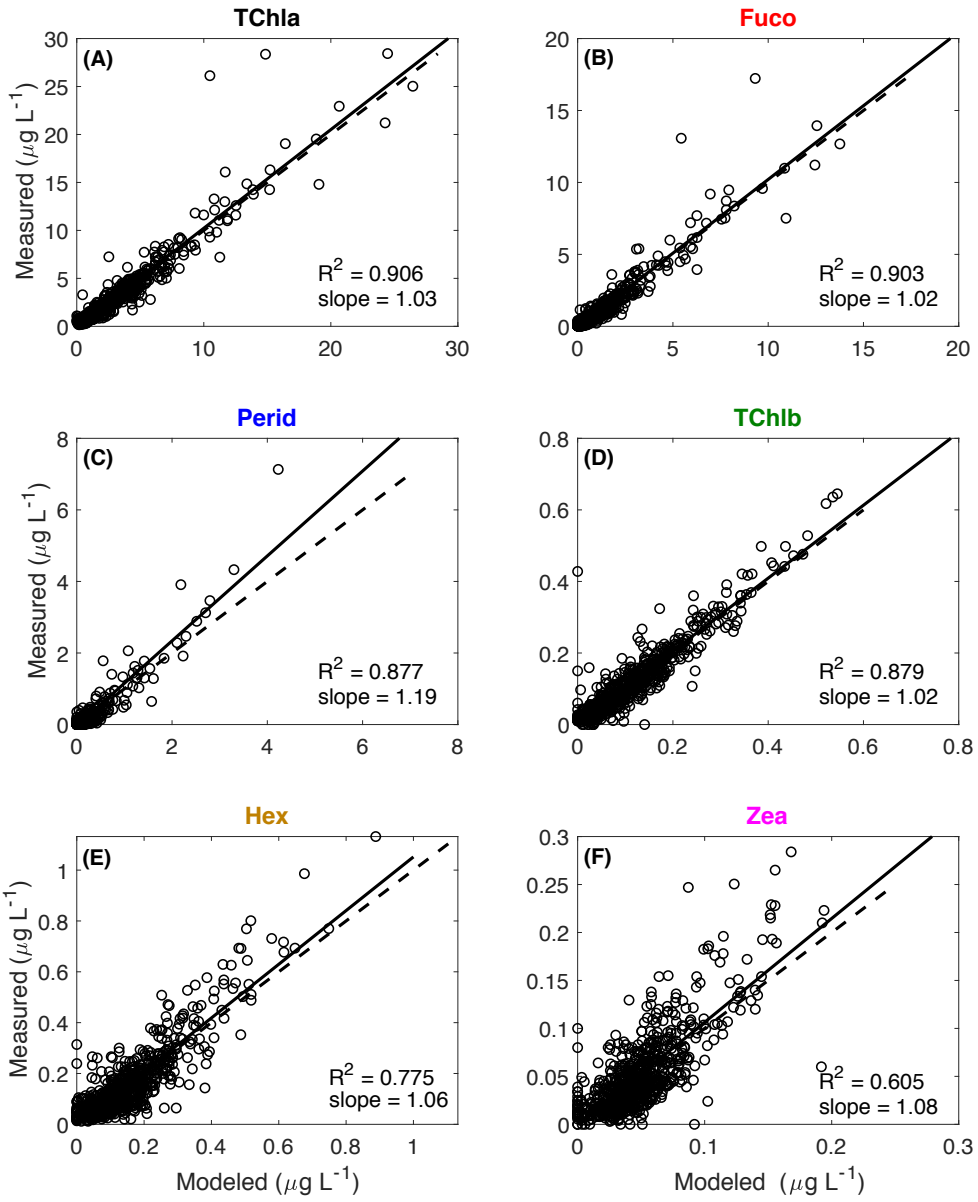

Figure S6. Model validations using all available HPLC observations from 2005-2018, including the model training data.

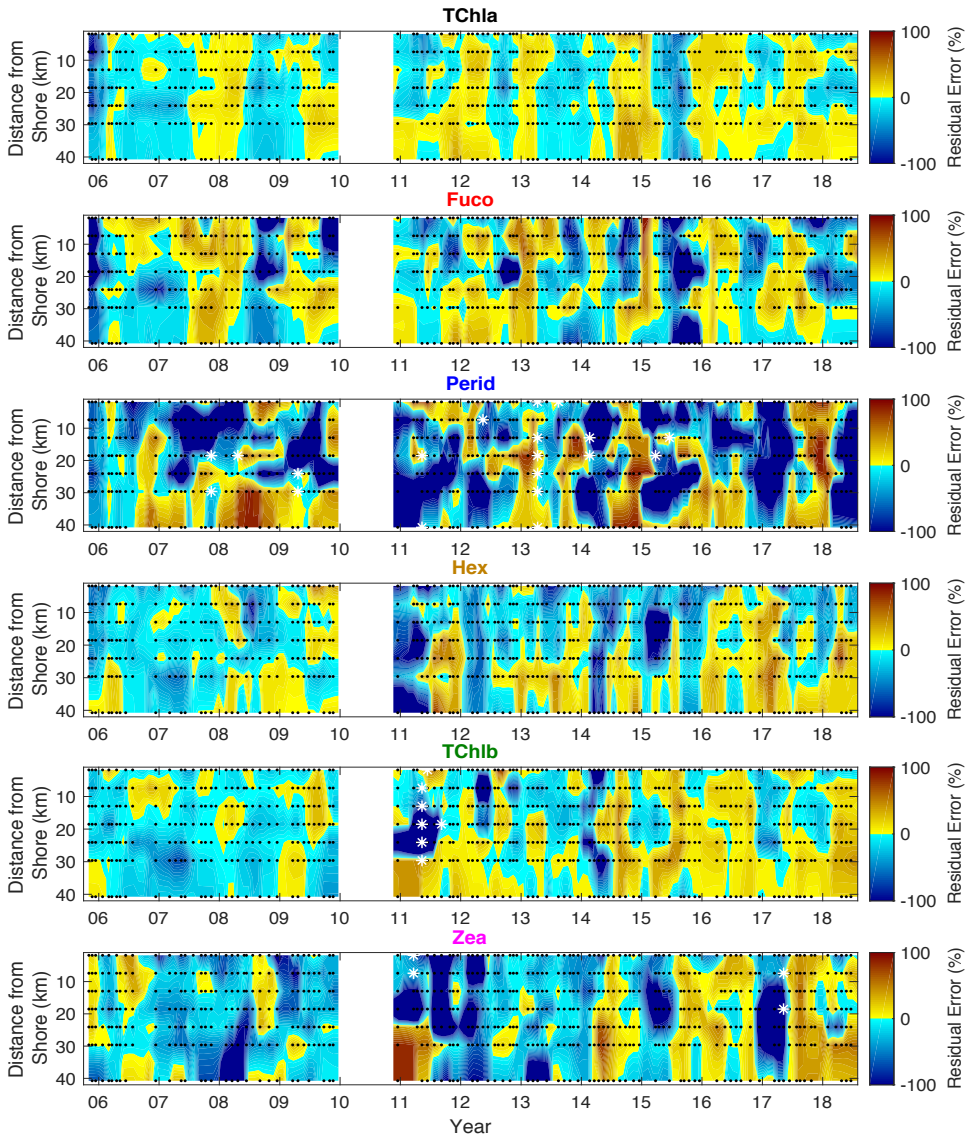

Figure S7. Spatiotemporal variations in residual error (%) in bio-optically modeled pigment concentrations with corresponding HPLC observations. The top of each panel corresponds to PnB station 1 on the mainland shelf. Positive values were never larger than 100%, and were 100% when the bio-optical models incorrectly predicted that a pigment concentration was below detection. We found 7, 19, and 4 occurrences where modeled TChlb, Perid, and Zea concentrations, respectively, had residual errors < -1000%. These large errors occur where HPLC pigment concentrations are low or below detection. The maximum absolute error where residual errors were < -1000% was 0.141, 0.271, and 0.092 μg L^{-1} for TChlb, Perid, and Zea concentrations, respectively, suggesting these large errors have minimal impact on the primary conclusions of this study. Ordinary kriging with an exponential-Bessel fitting model (GLOBEC Kriging Software Package v3.0) was used to smooth the data for this figure. Interpolation length scales are 30 days and 5 km in the cross-shelf direction. Black dots show where both HPLC and modeled pigment concentrations are available. White stars show residual errors < -1000%; these values are not considered in the kriging procedure.

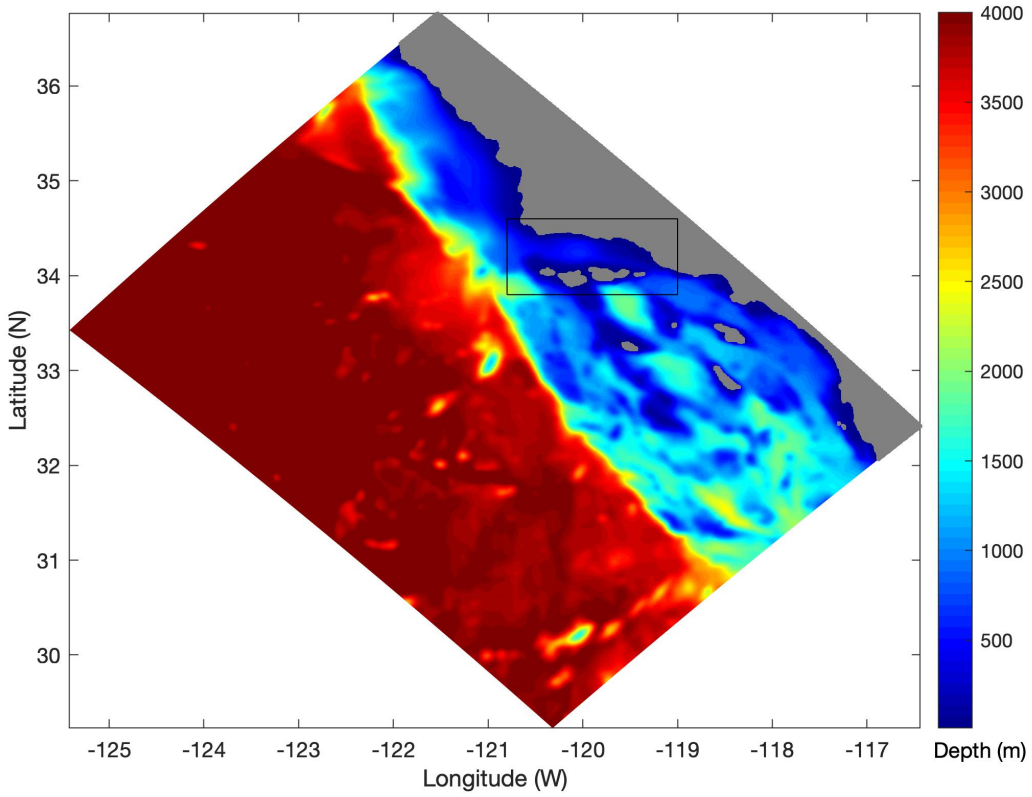

Figure S8. Domain of the Regional Ocean Modeling System used in the particle tracking models to determine SBC source waters. The boxed region roughly demarcates the Santa Barbara Channel.

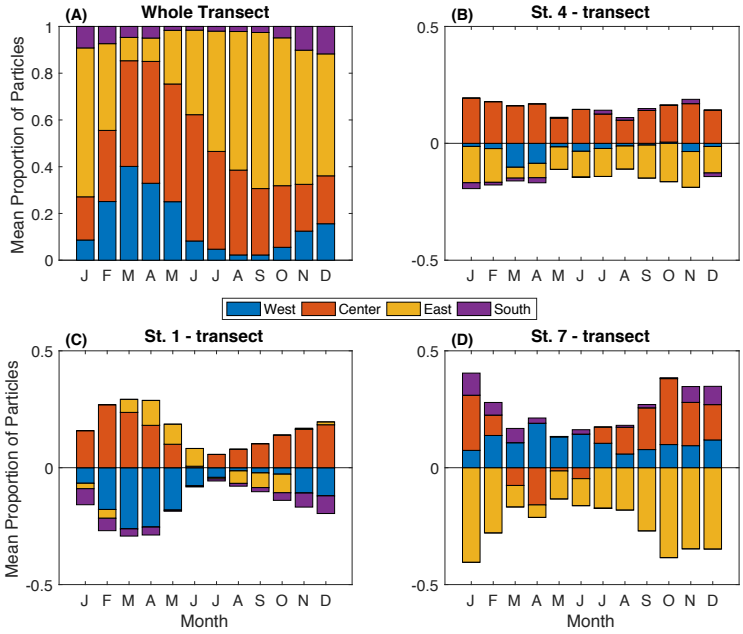

Figure S9. Mean annual cycle of the proportion of particles released from (A) all 34 release points on the PnB transect that originated from each of the four "origin boxes" shown in Figure 1 for 5-day advection times as determined by the Regional Ocean Modeling System particle tracking model. Panels B, C, and D show the difference in mean annual cycles from the four or five closest release points to PnB stations (B) 4, (C) 1, and (D) 7, from the transect mean annual cycle shown in (A). For results from 5- and 10-day advection times, see Figure 14 in the main text and Supporting Figure S8 below.

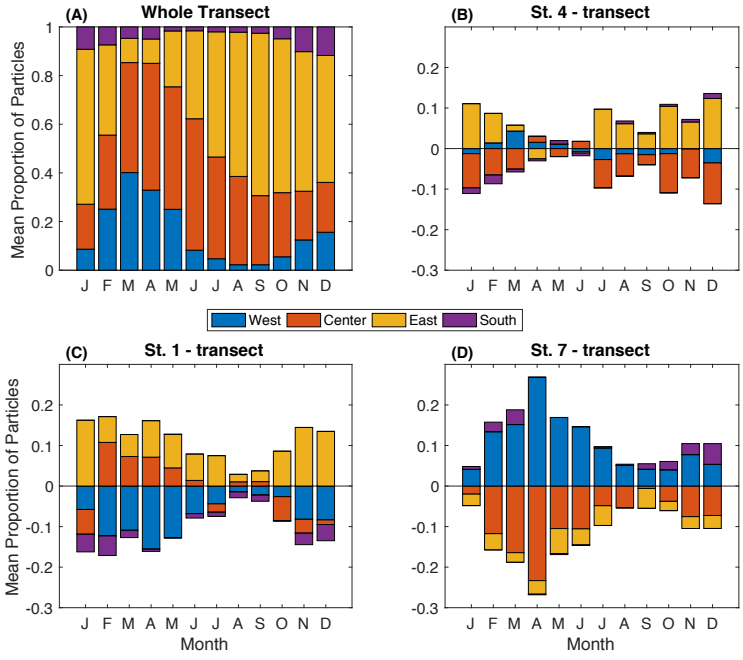

Figure S10. Same as Supporting Figure S7 for 15-day advection times.

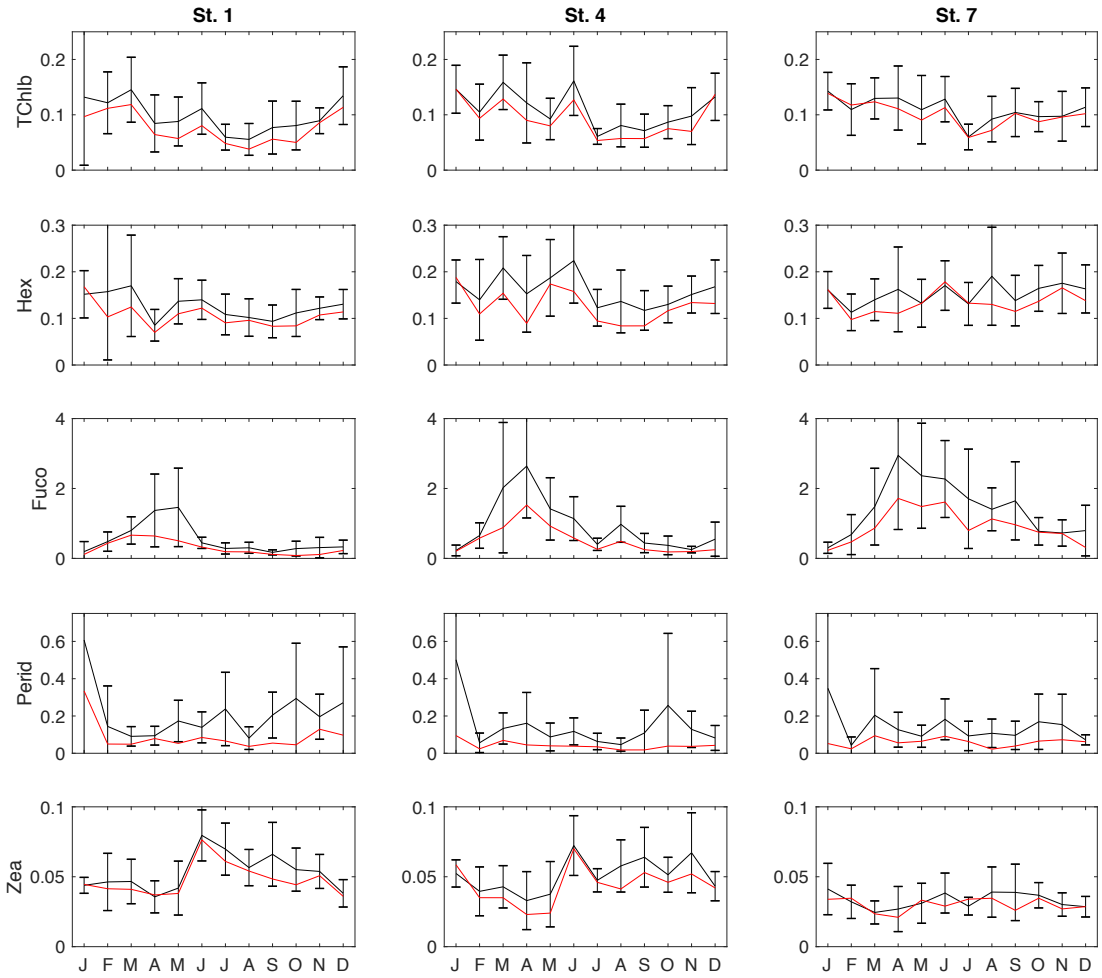

Figure S11. Mean +/- 95% confidence intervals (black) and median (red) annual cycles of the five marker pigment concentrations at Plumes and Blooms stations 1 (on the mainland continental shelf), 4 (in the central SBC), and 7 (on the Channel Islands continental shelf).

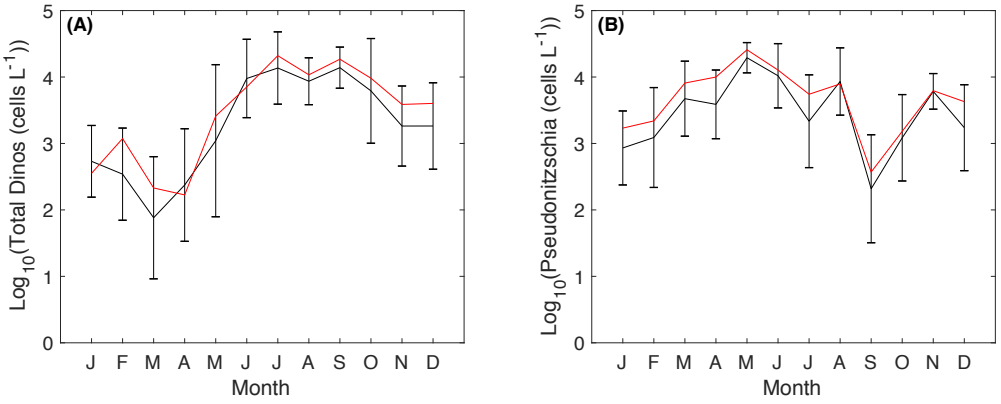

Figure S12. Mean +/- 95% confidence intervals (black) and median (red) annual cycles of (A) total dinoflagellates and (B) *Pseudonitzschia sp.* at Stearn's Wharf located on the mainland continental shelf in Santa Barbara, CA, as reported by the Southern California Coastal Ocean Observing System Harmful Algal Bloom monitoring project. The dinoflagellate species with available count data are *Akashiwo sanguinea*, *Alexandrium sp.*, *Dinophysis sp.*, *Lingulodinium polyedra*, *Prorocentrum sp.*, *Ceratium sp.*, and *Cochlodinium sp*. *Pseudonitzschia sp.* is the only reported diatom species. One cell mL^{-1} was added to counts of total dinoflagellates and total diatoms prior to log-transformation to prevent undefined values. Climatologies were computed from monthly mean time series determined from the weekly log-transformed values.

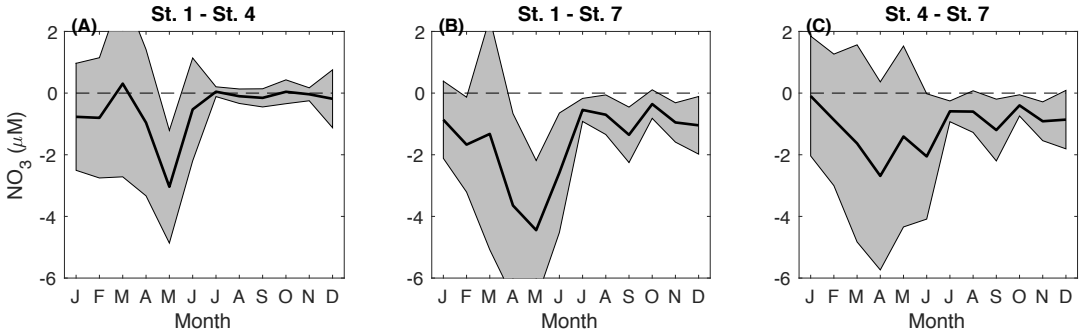

Figure S13. Spatial differences in mean annual cycles of nitrate concentrations across PnB stations (A) 1 and 4, (B) 1 and 7, and (C) 4 and 7. The shaded region around each line corresponds to 95% confidence intervals in the difference of each months mean nitrate concentration.

Table S1. Mean (standard deviation) goodness of fit statistics from the 500-fold model cross-validation procedure.

Pigment	R^2	RMSE	Median % error	Mean % error
TChla	0.87 (0.07)	1.03 (0.38)	17.2 (2.18)	26.3 (2.93)
TChlb	0.86 (0.04)	0.04 (0.004)	21.7 (2.51)	951 (968)
Hex	0.72 (0.06)	0.07 (0.01)	29.8 (3.69)	51.6 (6.26)
Fuco	0.87 (0.07)	0.54 (0.19)	35.0 (4.93)	62.7 (9.27)
Perid	0.88 (0.05)	0.13 (0.02)	98.8 (3.68)	2.05×10^3 (1.56×10^3)
Zea	0.54 (0.09)	0.03 (0.004)	38.3 (3.72)	461 (488)

Table S2. Long-term trend statistics for each pigment concentration monthly anomaly time series normalized to monthly means (see Figure 9 in the main text). Slopes were calculated according to the non-parametric Theil-Sen estimator, or as the median of all slopes calculated between all pairs of sequential observations. A Mann-Kendall test adjusted for autocorrelated time series following Hamed and Rao (1998) was used to calculate p-values. Statistically significant ($p < 0.05$) trends are in bold.

Pigment	Station	Slope (% yr^{-1})	p-value	N
Fluorometric chlorophyll *a* concentrations	1	-1.14	0.19	203
	2	-0.57	0.58	209
	3	-0.28	0.75	205
	4	0.54	0.51	208
	5	0.33	0.71	197
	6	0.50	0.61	190
	7	1.36	0.21	170
	Mean	0.36	0.71	212
TChla	1	-0.80	0.34	180
	2	-0.25	0.82	188
	3	-0.32	0.67	187
	4	0.81	0.29	191
	5	0.66	0.42	181
	6	0.88	0.35	172
	7	1.47	0.08	151
	Mean	1.33	0.18	206
Fuco	1	-1.70	0.08	180
	2	-1.12	0.11	188
	3	-0.83	0.22	187
	4	0.33	0.57	191
	5	0.04	0.95	181

	6	0.24	0.78	172
	7	1.00	0.25	151
	Mean	0.67	0.39	206
Perid	1	0.05	0.94	180
	2	0.11	0.84	188
	3	0.04	0.90	187
	4	0.46	0.58	191
	5	0.67	0.35	181
	6	0.45	0.56	172
	7	0.62	0.55	151
	Mean	0.93	0.51	206
Hex	1	-1.22	0.18	180
	2	**-1.75**	**0.03**	**188**
	3	**-2.08**	**0.01**	**187**
	4	-0.78	0.35	191
	5	-0.88	0.19	181
	6	**-1.53**	**0.03**	**172**
	7	-0.83	0.36	151
	Mean	-0.76	0.35	206
TChlb	1	-0.15	0.83	180
	2	-1.16	0.29	188
	3	-1.40	0.18	187
	4	-0.15	0.87	191
	5	-0.59	0.58	181
	6	-0.57	0.53	172
	7	0.46	0.66	151
	Mean	-0.16	0.86	206
Zea	1	0.09	0.88	180
	2	-0.34	0.52	188
	3	-0.16	0.80	187
	4	-0.60	0.39	191
	5	-0.38	0.55	181
	6	-0.96	0.14	172
	7	-1.03	0.16	151
	Mean	-0.09	0.88	206

Appendix 3. Supplemental Information for Chapter IV

Supporting File S1. Compilation of protistan lineages and corresponding trophic modes based on several refereed and non-refereed sources (see Chapter IV).

Supporting Figures and Tables

Table S1. Ensemble taxonomic assignments and BLASTN statistics of putative Fuco- and Perid-containing dinoflagellate ASVs.

ASV ID	Putative biomarker	ASV taxonomic annotation[a]	Genus of closest hit[b]	Query cover (%)	Identity (%)	References[c]
sv8	Fuco	*Dinophyceae*	*Karenia*, *Ensiculifera*, *Pentapharsodinium*, *Shimiella*, *Gymnodinium*	100	100	1, 2, 3
sv327	Fuco	*Gymnodinium*	*Gymnodinium*	100	100	2, 4
sv404	Fuco	*Protodinium*	*Protodinium*	100	100	2, 4
sv113	Perid	*Tripos*	*Tripos*	100	96.8	1, 5
sv2442	Perid	*Tripos*	*Tripos*	100	96.03	1, 5

[a]Genus assignments determined with the ensemble taxonomic assignment procedure (see main text) are provided where available, otherwise the taxonomy assigned at the lowest annotated rank

[b]Where multiple best hits were found for an ASV, all genera are listed. BLAST searches excluded sequences from uncultured organisms and environmental samples.

[c]References: (1) Zapata et al. 2012; (2) Carreto et al. 2001; (3) Ok et al. 2020; (4) Guiry and Guiry, 2021; (5) Gomez et al., 2013.

Supporting Table S2. Same as Table 1 for additional multiple linear regression models to support the interpretations provided in the text. Predictor variable abbreviations: MLD, mixed layer depth; PAR, photosynthetically available radiation; Fuco-dino, summed POC or RSA of three putative Fuco-containing dinoflagellate ASVs (see main text, Supp. Table 1); Perid-dino, summed POC or RSA of three putative Perid-containing dinoflagellate ASVs (see main text, Supp. Table 1); Chloro-eco1, Chloro-eco2, Chloro-eco3, summed POC or RSA of all ASVs included in *Chlorophyta* ecotypes 1, 2, or 3, respectively (see main text, Supp. Figs. 7-8); sv15-*P. globosa*, the POC or RSA of the putative *Phaeocystis globosa* ASV associated with underestimation of prymnesiophyte biomass contributions with HPLC pigment methods (see main text, Supp. Fig. 10).

		R^2	N	MLD (p-val)	PAR (p-val)	Fuco-dino* (p-val)	Perid-dino* (p-val)	Chloro-eco1* (p-val)	Chloro-eco2* (p-val)	Chloro-eco3* (p-val)	sv15-*P. globosa** (p-val)
Diatom	PG	0.34	183	-0.27 (< 0.001)	-0.11 (0.12)	-0.47 (< 0.001)					
	PCC	0.37	189	0.12 (0.050)		-0.62 (< 0.001)					
	PCC	0.40	189	0.032 (0.62)	-0.21 (0.002)	-0.53 (< 0.001)					
Dino	PG	0.75	183	-0.17 (< 0.001)	-0.0003 (0.99)	0.72 (< 0.001)	-0.39 (< 0.001)				
	PCC	0.57	189	-0.092 (0.063)		0.67 (< 0.001)	-0.29 (< 0.001)				
	PCC	0.57	189	-0.069 (0.21)	0.056 (0.33)	0.64 (< 0.001)	-0.27 (< 0.001)				
Chloro	PG	0.56	183	-0.12 (0.025)	0.11 (0.054)			0.58 (< 0.001)	0.36 (< 0.001)	-0.37 (< 0.001)	
	PCC	0.54	189	0.16 (0.002)				0.70 (< 0.001)	0.17 (0.003)	-0.18 (0.001)	
	PCC	0.54	189	0.17 (0.002)	0.042 (0.51)			0.69 (< 0.001)	0.19 (0.003)	-0.18 (0.003)	
Prym	PG	0.56	183	-0.21 (< 0.001)							0.73 (< 0.001)
	PCC	0.57	189	-0.024 (0.62)							0.76 (< 0.001)
	PCC	0.73	189	0.12 (0.005)	0.43 (< 0.001)						0.72 (< 0.001)

*POC concentrations were used to predict PG POC residuals, while RSAs were used to predict PCC RSA residuals.

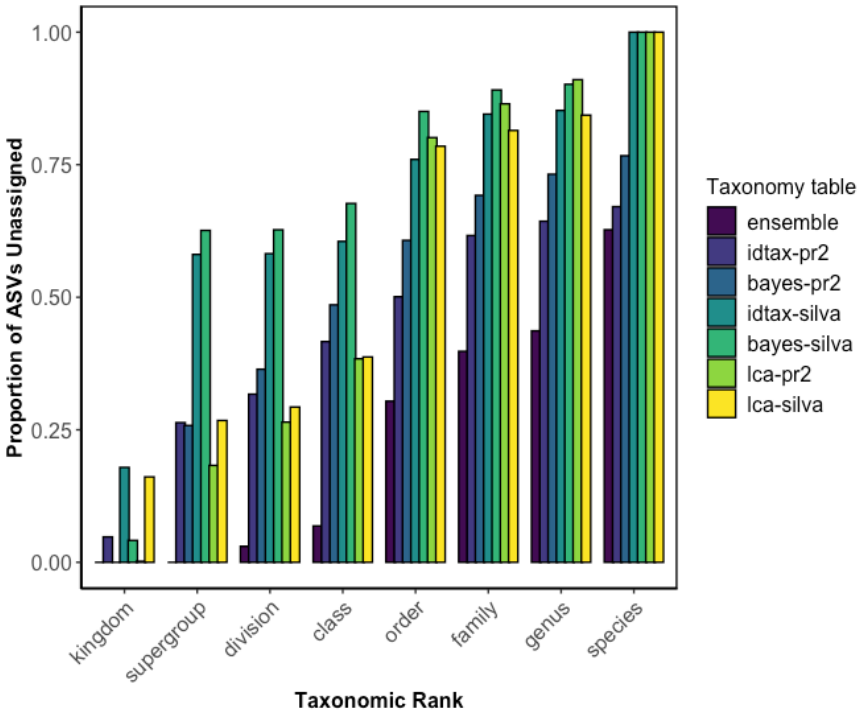

Figure S1. Comparison of the proportion of protistan ASVs that were left unassigned at each taxonomic rank by ensemble and individual taxonomic assignment methods.

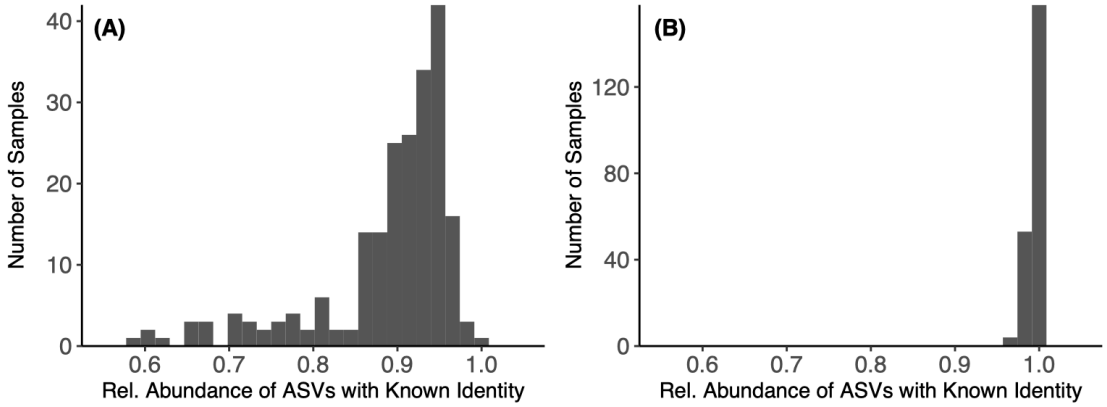

Figure S2. Distributions of protistan relative sequence abundances of ASVs that were unambiguously assigned as phytoplankton or non-phytoplankton in surface ocean samples (A) following the initial phytoplankton classification procedure described in the main text and (B) subsequently assigning the remaining *Dinophyceae*, *Cryptophyta*, and *Haptophyta* ASVs with an unknown phytoplankton classification as phytoplankton.

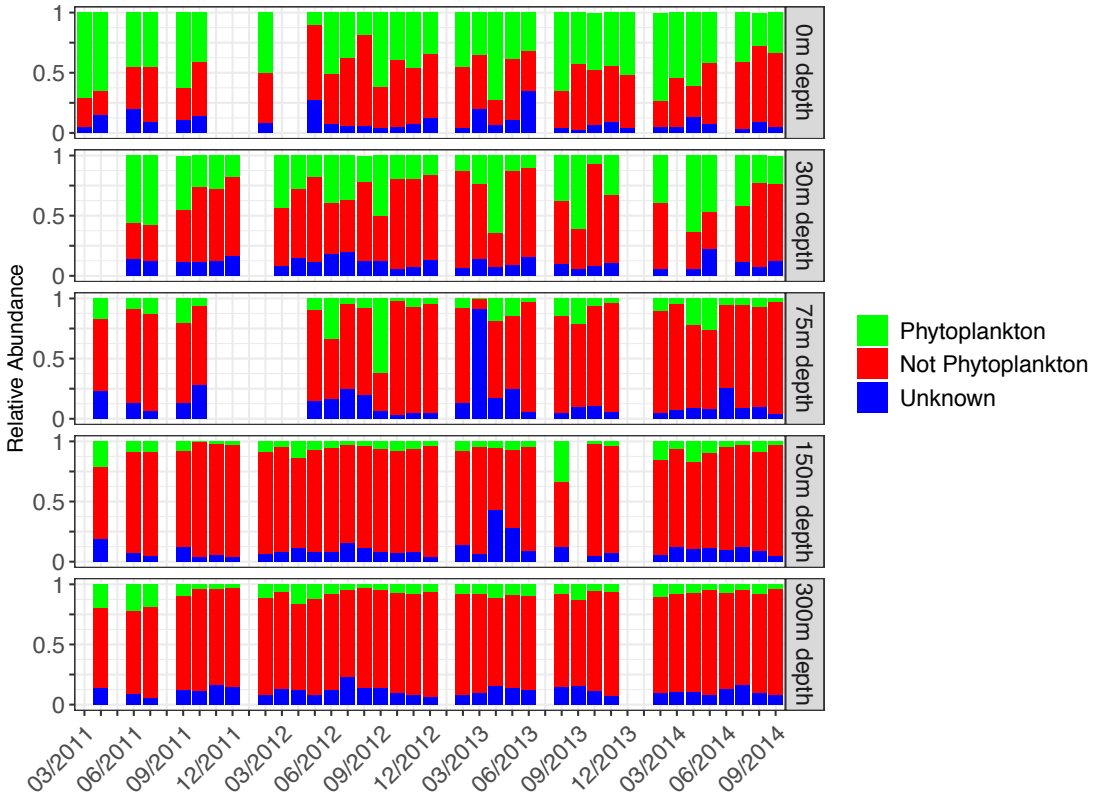

Figure S3. Vertical profiles of the relative sequence abundances of protistan ASVs that were classified as phytoplankton (green), non-phytoplankton (red), or unknown (blue) following the initial phytoplankton classification procedure described in the main text.

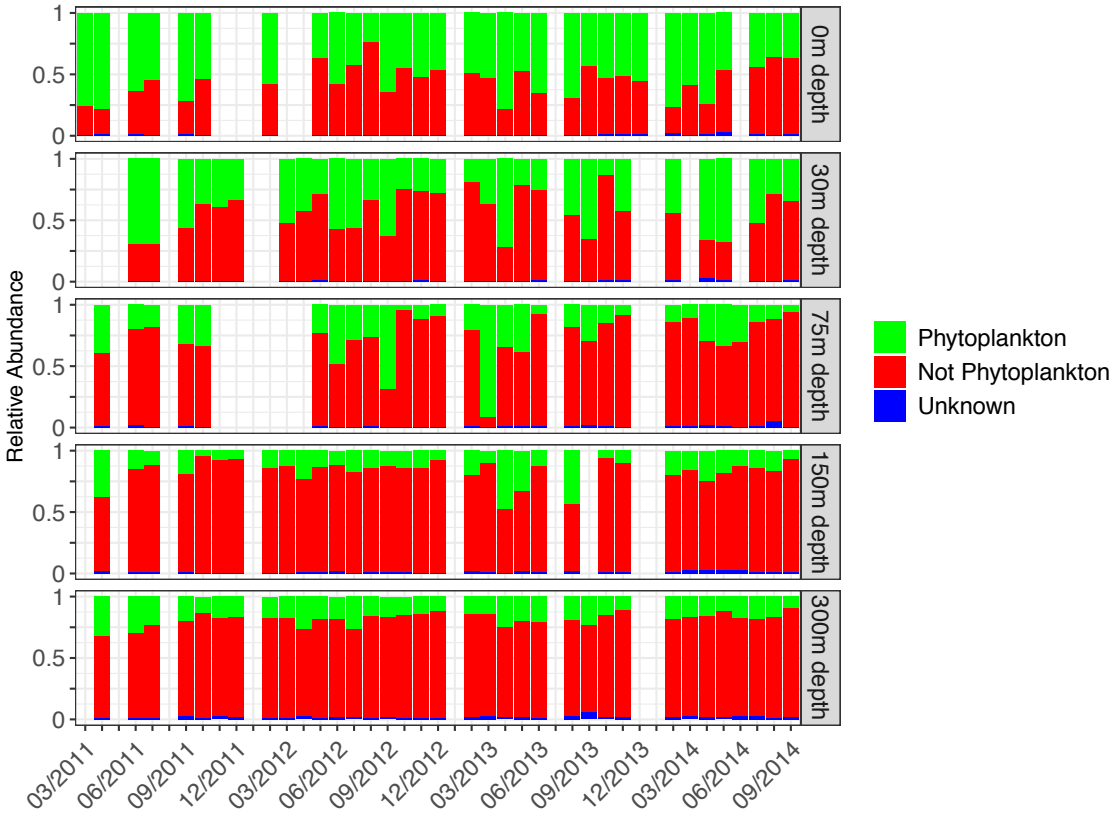

Figure S4. Vertical profiles of the relative sequence abundances of protistan ASVs that were classified as phytoplankton (green), not phytoplankton (red), or unknown (blue) following the initial phytoplankton classification procedure described in the main text and subsequently assigning the remaining *Dinophyceae*, *Cryptophyta*, and *Haptophyta* ASVs with an unknown phytoplankton classification as phytoplankton.

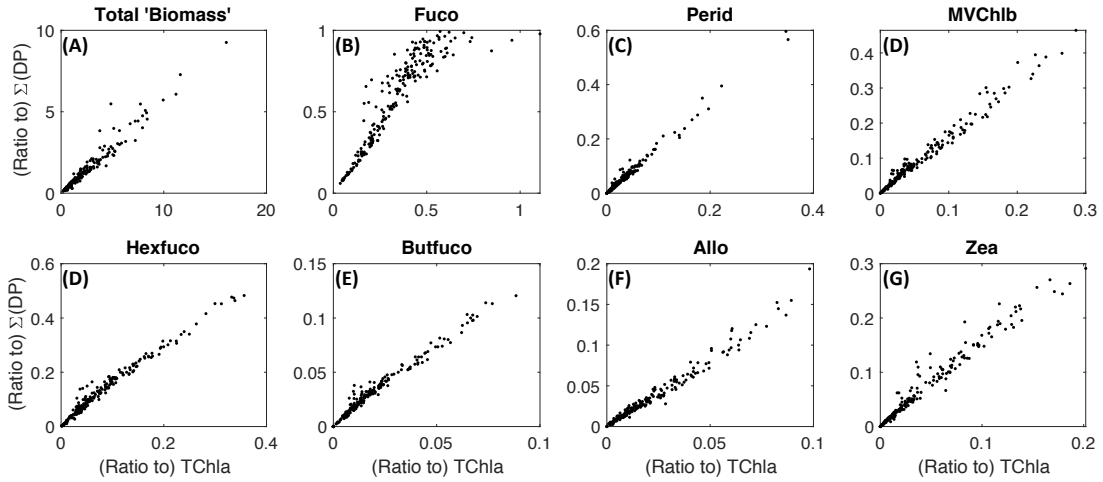

Figure S5. Relationships between (A) total pigment biomass (total chlorophyll *a* and the sum of the seven diagnostic pigment concentrations as in Vidussi et al., 2001) and (B-G) phytoplankton pigment biomass contributions (biomarker pigment ratios to total chlorophyll *a* or the sum of the seven diagnostic pigment concentrations as in Vidussi et al., 2001).

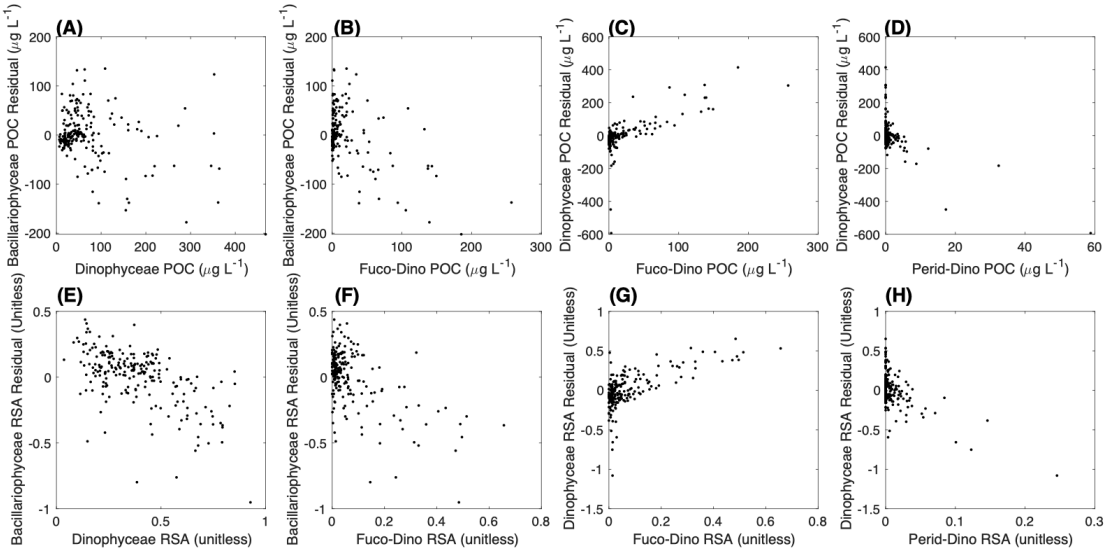

Figure S6. Inter- and intra-lineage predictors of uncertainty in diatom and dinoflagellate biomass contribution estimates. (A, E) show relationships between total *Dinophyceae* POC or RSA and diatom POC or RSA residuals, (B, C, F, G) show relationships between the total POC or RSA of three putative Fuco-containing dinoflagellate ASVs (see main text and Supporting Table 1) and (B, F) diatom or (C, G) dinoflagellate POC or RSA residuals, and (D, H) show relationships between the total POC or RSA of two putative Perid-containing dinoflagellate ASVs (see main text and Supporting Table 1) and dinoflagellate POC or RSA residuals.

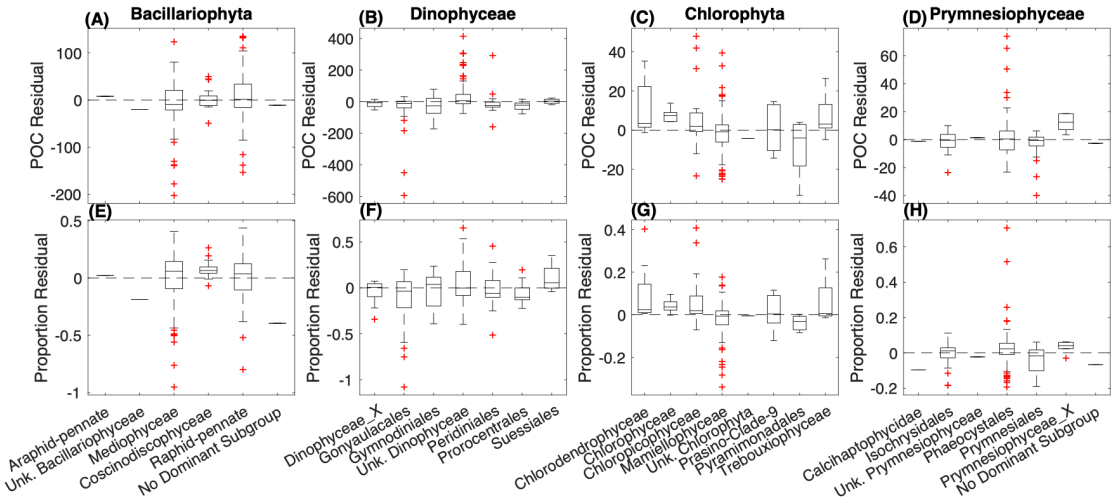

Figure S7. Assessment of potential intra-group variations in (A-D) POC and (E-H) RSA residual value distributions for each of the four dominant phytoplankton groups. Shown are the distributions of residual values across groups of samples that are dominated by a particular class (for *Chlorophyta*) or order (for *Bacillariophyta*, *Dinophyceae*, and *Prymnesiophyceae*). Single horizontal lines indicate the given taxonomic group was only dominant in one sample.

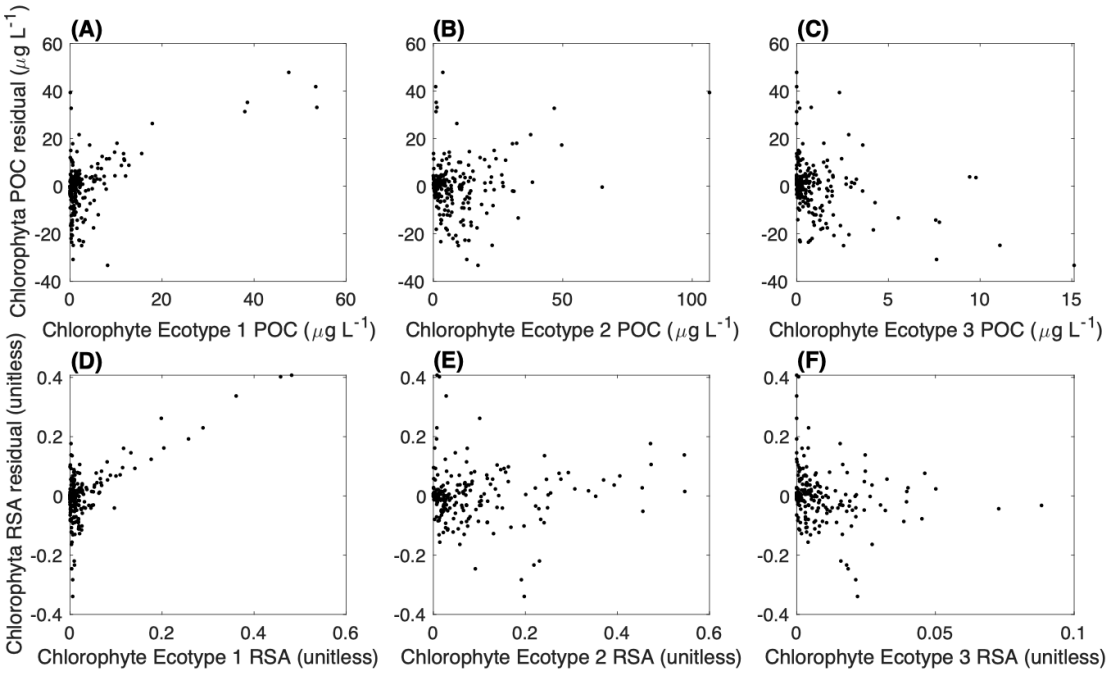

Figure S8. Relationships between the total (A-C) POC concentrations and (D-F) relative sequence abundances of each of the three *Chlorophyta* ecotypes and *Chlorophyta* POC and relative sequence abundance residuals.

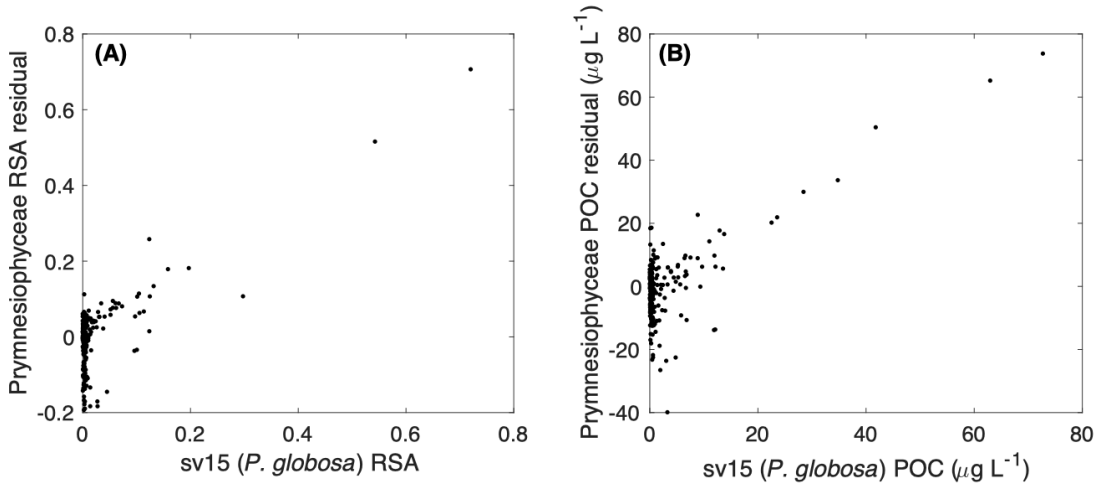

Figure S9. Relationships between the putative *Phaeocystis globosa* ASV's (sv15) (A) relative sequence abundance and (B) POC concentration and *Prymnesiophyceae* relative sequence abundance and POC residuals.

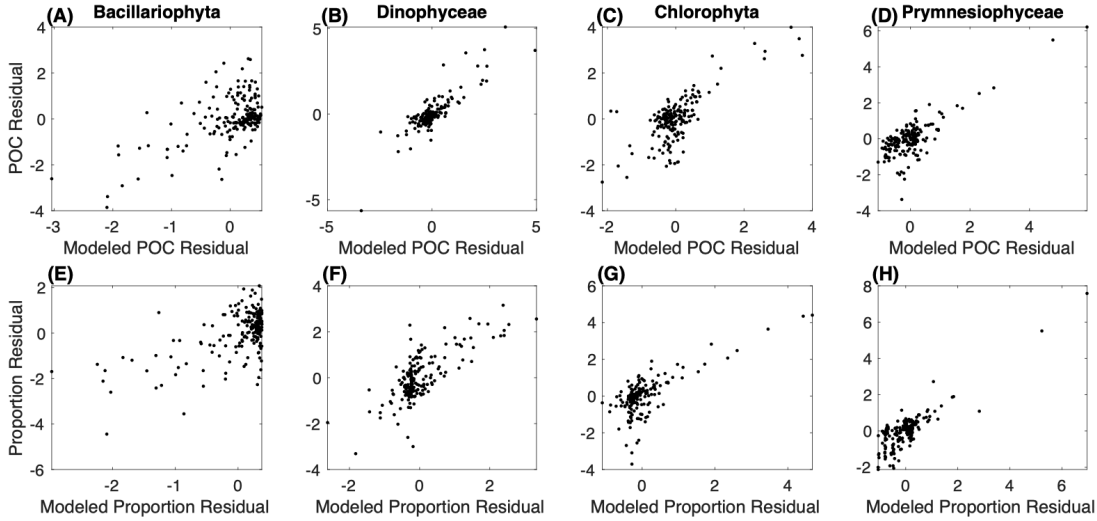

Figure S10. Multiple linear regression model fits for each of the four phytoplankton group's (A-D) POC and (E-H) relative sequence abundance residuals. All variables were z-scored prior to multiple linear regression analysis. Therefore all values have standard deviation units.

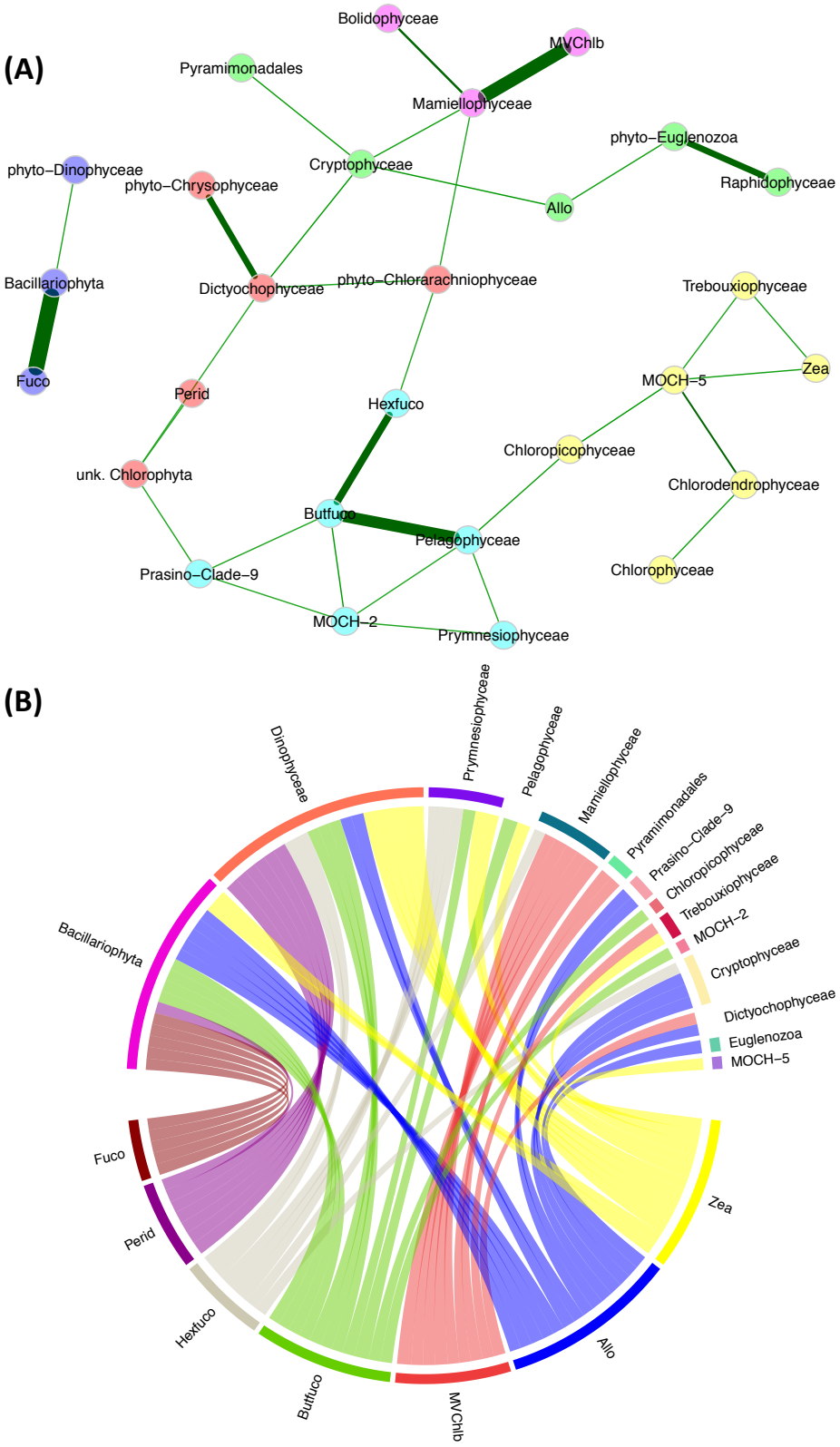

Figure S11. Same as Figure 3, but for compositional subnetworks including only (A) classes or (B) ASVs that were assigned as phytoplankton alongside biomarker pigment ratios to total chlorophyll *a*.

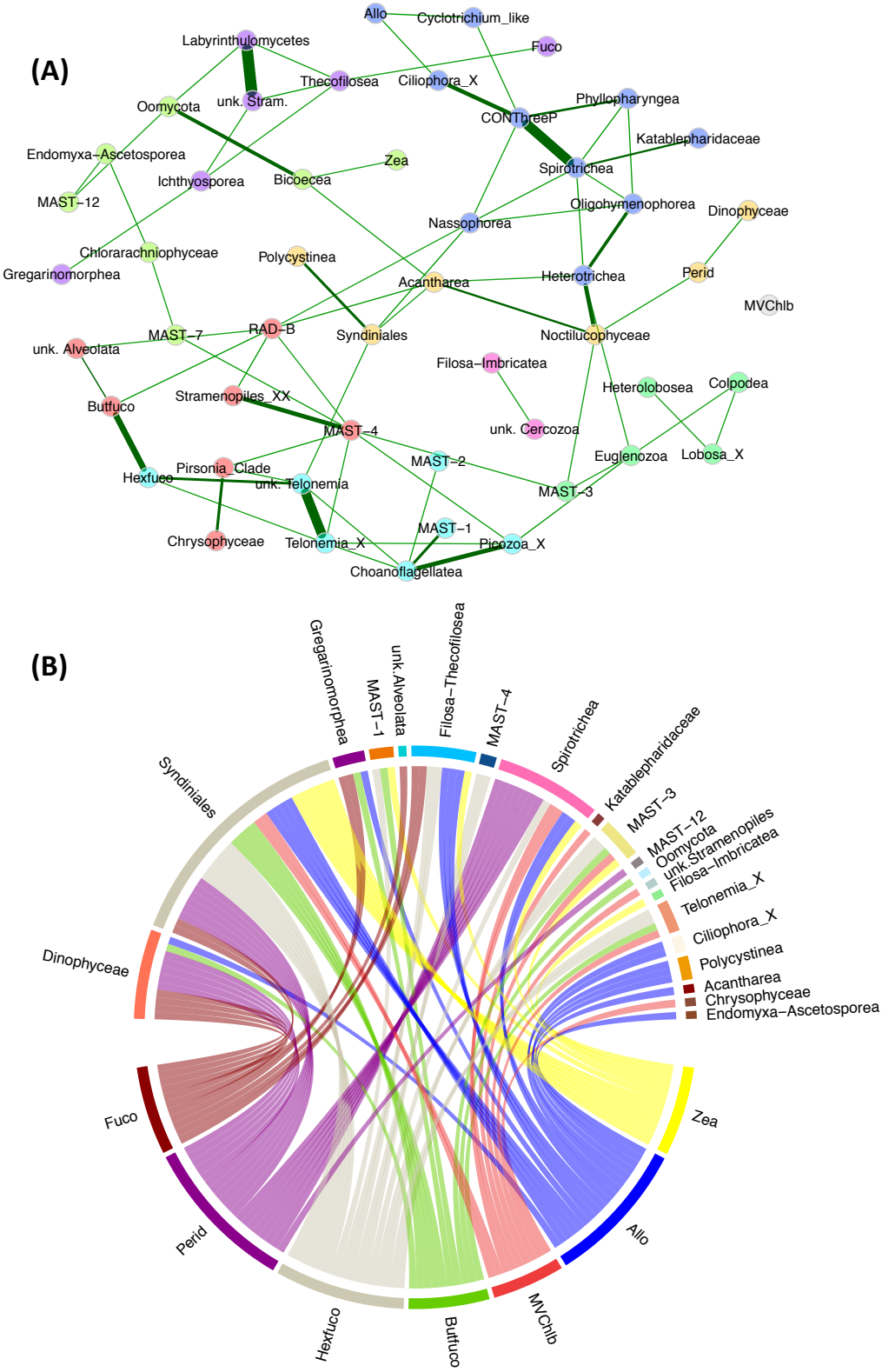

Figure S12. Same as Figure 4, but for compositional networks including both phytoplankton and other protistan (A) classes or (B) ASVs with biomarker pigment ratios to total chlorophyll *a*.

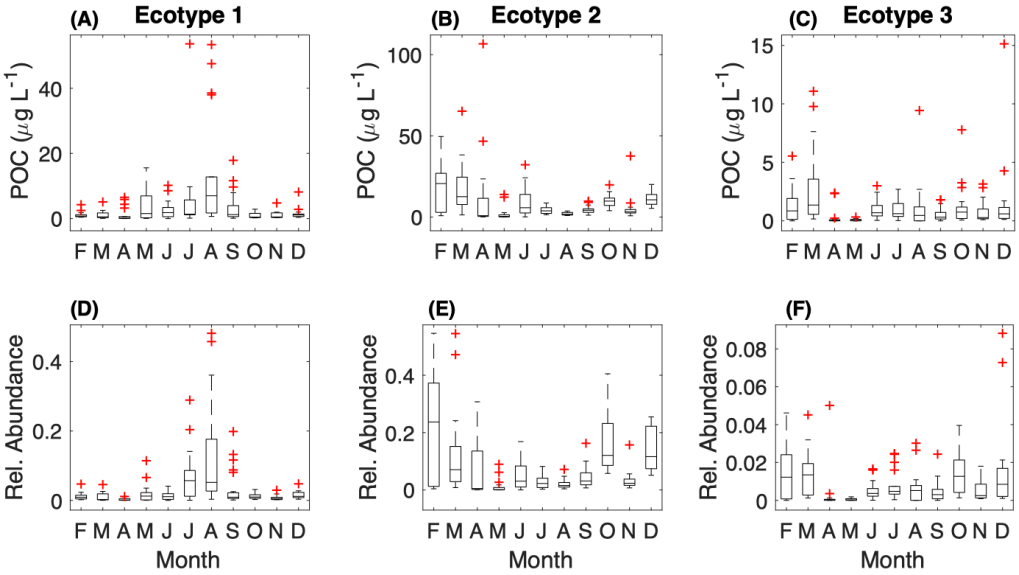

Figure S13. Distributions of (A-C) POC concentrations and (D-F) relative sequence abundances of the three *Chlorophyta* ecotypes across samples separated according to the month sampled.