LIMNOLOGY and OCEANOGRAPHY

ASLO

Limnol. Oceanogr. 68, 2023, 361–376

doi: 10.1002/lno.12274

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

Dylan Catlett,[1*] **David A. Siegel**,[1,2,3] **Paul G. Matson**,[3,4] **Emma K. Wear**,[3,5] **Craig A. Carlson**,[3,5] **Thomas S. Lankiewicz**,[5] **M. Debora Iglesias-Rodriguez**[3,5]

[1]Earth Research Institute, UC Santa Barbara, Santa Barbara, California
[2]Department of Geography, UC Santa Barbara, Santa Barbara, California
[3]Marine Science Institute, UC Santa Barbara, Santa Barbara, California
[4]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee
[5]Department of Ecology, Evolution, and Marine Biology, Santa Barbara, California

Abstract

Quantifying phytoplankton composition is critical to predicting marine ecosystem structure and function. DNA meta-barcoding and high-performance liquid chromatography (HPLC) pigment analysis are two widely used methods for assessing phytoplankton composition; however, comparing their performance has been done only rarely. Here, we integrate DNA meta-barcoding and HPLC pigment observations to determine eukaryotic phytoplankton composition in the Santa Barbara Channel, California. We find that both methods identify the same four dominant eukaryotic phytoplankton taxa (diatoms, dinoflagellates, chlorophytes, and prymnesiophytes), but inter- and intra-lineage variability in biomarker pigmentation (associated with both a lack of taxonomic specificity of biomarker pigments and intrinsic differences in accessory pigmentation) drives substantial disagreement between the methods. Covariation network analysis circumvents this disagreement and reveals that diverse assemblages of phytoplankton and other protists covary with distinct suites of biomarker pigments. Our results highlight the strengths and weaknesses of each method in characterizing phytoplankton composition and reveal novel insights into phytoplankton physiology that could only be gained by integrating the two methods. Finally, we suggest a path to monitor eukaryotic plankton communities on unprecedented spatiotemporal scales based on the covariation of unique phytoplankton and protistan assemblages with remotely sensible phytoplankton pigment concentrations.

Phytoplankton production fuels marine food webs and the biological carbon pump (Ryther 1969; Guidi et al. 2016). Phytoplankton composition determines the efficiency of the biological carbon pump and of the transfer of phytoplankton production to higher trophic levels, dictating marine ecosystem structure and function (Guidi et al. 2016; Lin et al. 2017). Thus, quantifying phytoplankton composition (here, meaning relative abundances or biomass contributions) and/or concentrations of phytoplankton taxa represents a critical step in efforts to understand and predict marine ecosystem structure and function.

Quantifying phytoplankton composition is difficult due to the high diversity of phytoplankton (De Vargas et al. 2015). Methods available for quantifying phytoplankton composition include high-performance liquid chromatography (HPLC) pigment analysis, amplicon sequencing of DNA "barcode" genes (DNA meta-barcoding), other "meta-omics" techniques (meta-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 phytoplankton composition have been proposed but rely on one of these methods for formulation and validation (Uitz et al. 2015; Chase et al. 2017). Both HPLC pigment and DNA meta-barcoding methods entail relatively efficient sample analysis procedures and are widely applied to assess phytoplankton composition

*Correspondence: dylan.catlett@whoi.edu, dsc@ucsb.edu

Additional Supporting Information may be found in the online version of this article.

Author Contribution Statement: D.C., D.A.S., P.G.M., E.K.W., C.A.C., M.D.I.R. designed the research. D.C., P.G.M., E.K.W., T.S.L. collected and analyzed samples. D.C. analyzed the data. D.C. wrote the manuscript with contributions from all coauthors.

(Lima-Mendez et al. 2015; Kramer and Siegel 2019). Recent work suggests that HPLC and DNA meta-barcoding analyses provide estimates of phytoplankton composition that are more comparable to one another than to other methods, in part because they both allow sampling of a broad range of size classes (Gong et al. 2020).

DNA meta-barcoding analysis results in a collection of amplicon sequence variants (ASVs), which serve as proxies for "species" in microbial ecology applications (Callahan et al. 2016). Although the total sequence counts recovered from a sample or sequencing run is constrained by sample preparation and analysis procedures (Gloor et al. 2017), ASV relative sequence abundances are often estimated by normalizing each ASV's sequence counts to the total counts recovered from each sample (De Vargas et al. 2015; Berdjeb et al. 2018). Interpreting compositional (relative) data are difficult because variability in a single ASV's relative abundance 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). Analytical uncertainty in DNA meta-barcoding estimates of phytoplankton composition is difficult to evaluate (Bradley et al. 2016; Wear et al. 2018). However, recent work suggests rigorously evaluated DNA meta-barcoding workflows provide reasonably accurate and precise estimates of phytoplankton composition (Catlett et al. 2020*a*; Yeh et al. 2021). Overall, these data offer high taxonomic resolution of phytoplankton composition and are increasingly used to assess the roles of phytoplankton communities in marine ecosystems and biogeochemical cycles (Guidi et al. 2016; Lin et al. 2017).

HPLC analysis quantifies the concentrations of a suite of phytoplankton pigments, some of which are presumed to serve as biomarkers for particular phytoplankton taxa (Mackey et al. 1996; Jeffrey et al. 2011; Kramer and Siegel 2019). The primary weaknesses in HPLC pigment analysis are (1) limited taxonomic resolution (to approximately the class level; Mackey et al. 1996; Kramer and Siegel 2019); (2) variability in pigment concentrations independent from phytoplankton biomass due to physiological status and other factors (Goericke and Montoya 1998; Schlüter et al. 2000); and (3) lack of specificity of commonly used biomarker pigments (Jeffrey et al. 2011). Despite these drawbacks, HPLC sample analysis methods are rigorously evaluated and standardized (Van Heukelem and Thomas 2001), and direct links between phytoplankton pigments and bio-optical properties provide a path to observe phytoplankton composition from satellite ocean color (Bricaud et al. 2004; Chase et al. 2017). Biomarker pigment concentrations can be transformed to compositions by normalizing to the total chlorophyll *a* (TChla) concentration or the sum of all biomarker pigment concentrations (Mackey et al. 1996; Vidussi et al. 2001). Alternatively, phytoplankton pigment "communities" can be identified from pigment data based on the covariation of pigment concentrations 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 assemblages, bio-optically modeled with relatively high accuracy, and analyzed further to determine oceanographic forcings of phytoplankton taxa on large spatiotemporal scales (Catlett et al. 2021*a*). These analyses are similar to those employed in recent studies that take a systems-level approach to characterize covariation among microbial ASVs using covariation networks (Lima-Mendez et al. 2015; Berdjeb et al. 2018).

Given the numerous approaches for quantifying phytoplankton composition, research is needed to develop coherent approaches for characterizing phytoplankton composition across scales of space, time, and phytoplankton diversity. Here, we integrate HPLC pigment and DNA meta-barcoding phytoplankton composition and concentration estimates using a large data set of concurrent observations from the Santa Barbara Channel, California. Our results highlight the strengths, weaknesses, and assumptions inherent in estimating phytoplankton composition and concentrations for each method and demonstrate that integrating these methods provides novel insights. Finally, we suggest that characterizing the plankton assemblages that covary with remotely sensible phytoplankton pigments may offer a path to monitor eukaryotic plankton assemblages on regional to global scales via satellite ocean color observations.

Methods

Overview of study site, sampling, and data availability

The Santa Barbara Channel (SBC) is a productive marine ecosystem at the boundary of the California Current System and Southern California Bight. Variability in SBC oceanographic properties is dominated by the annual wind-driven upwelling cycle, which is in turn modulated by remotely forced climate oscillations (Catlett et al. 2021*a*). Diatoms dominate the accumulation of phytoplankton biomass associated with spring-time upwelling in the SBC, though other phytoplankton also respond positively to upwelling-induced nutrient enrichment of the surface ocean (Taylor et al. 2015; Catlett et al. 2021*a*). High concentrations of other phytoplankton, including dinoflagellates (Catlett et al. 2021*a*), prymnesiophytes (Goodman et al. 2012), and chlorophytes (Countway and Caron 2006), have also been observed in and around the SBC.

The Plumes and Blooms project has sampled seven stations on a cross-SBC transect approximately monthly since August 1996 (Catlett et al. 2021*a*). Here, we consider data obtained from cruises conducted between March 2011 and September 2014 where concurrent observations of HPLC pigment concentrations and DNA meta-barcoding of the V9 hypervariable region of the 18*S small subunit rRNA gene* (henceforth, 18S rDNA) are available. Most of our analysis is focused on near-surface samples (nominal depth 1 m), though we also consider DNA meta-barcoding samples collected at depths of 30, 75,

150, and 300 m from Sta. 4 in the center of the transect (only surface samples are collected from the other stations). We supplement our analysis with Plumes and Blooms oceanographic observations including mixed layer depth and particulate organic carbon (POC) concentrations, and satellite observations of photosynthetically available radiation (Supporting Information Text S1). Methods used for sampling and analysis of pigment and oceanographic data are available elsewhere (Catlett et al. 2021*a*), described briefly in Supporting Information Text S1, and data are available through the Environmental Data Initiative (Catlett et al. 2020*b*). Amplicon sequencing and bioinformatic analysis follow methods described previously (Catlett et al. 2020*a*) and are detailed in Supporting Information Text S1. Raw sequence data are available in the National Center for Biotechnology Information's Sequence Read Archive (accession number PRJNA532583), and curated data are available through the Environmental Data Initiative (Catlett et al. 2022).

DNA meta-barcoding taxonomic assignments and data preprocessing

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. 2020*a* and Supporting Information Fig. S1). We implemented an ensemble taxonomic assignment approach (ensembleTax R package v1.1.1; Catlett et al. 2021*b*) with the goal of increasing the specificity of taxonomic assignments for ASVs in our data set. Initial taxonomic assignments were predicted with three widely used algorithms (Altschul et al. 1990; Huson et al. 2007; Wang et al. 2007; Murali et al. 2018) and both the Protistan Ribosomal Reference database (v4.12.0; Guillou et al. 2012) and the Silva SSU reference database (v138; Quast et al. 2012). Detailed descriptions of methods used for ensemble assignments and data preprocessing are available in Supporting Information Text S2.

All analyses presented here rely on a data set comprising 13,308 protistan ASVs derived from 345 discrete seawater samples and use the Protistan Ribosomal Reference database taxonomic nomenclature. Sequencing depth ranged from 11,804 to 225,911 protistan sequence reads per sample. Sequence counts of each protistan ASV were normalized to the total protistan sequence counts within each sample to determine ASV relative sequence abundances. Where duplicate or triplicate samples were available, mean relative sequence abundances were computed. Most analyses consider a subset of 215 surface samples, in which 6568 total protistan ASVs were detected.

Identification of phytoplankton ASVs

Phytoplankton ASVs must be identified accurately as their misidentification will bias estimates of phytoplankton composition since the abundance of each ASV is dependent on the abundances of all other ASVs in the composition. This is complicated by the growing recognition of mixotrophy as a widespread trophic strategy in marine protists (Mitra et al. 2016). Here, phytoplankton are defined as protistan taxa that are thought to include only photoautotrophic and/or constitutive mixotrophic representatives (possessing an inherent capacity to photosynthesize). We compiled a collection of taxonomic names with corresponding trophic modes following the definitions of Mitra et al. (2016) using several recent literature compilations and searches of other refereed and non-refereed sources and identified phytoplankton ASVs using their predicted taxonomic assignments (Supporting Information Text S3; Files S1, S2; Figs. S2–S4). Notably, ASVs with taxonomy assigned as unknown Dinophyceae, Cryptophyta, or Haptophyta could not be unambiguously classified but are assumed to be phytoplankton in the present analysis (Supporting Information Text S3).

Parallel analysis of concentrations and composition

Because HPLC pigment data are concentrations and DNA meta-barcoding data are compositions, transforming one or both data sets is required prior to integrating them. Methods for constraining pigment concentration data to compositions are imperfect (Irigoien et al. 2004; Catlett and Siegel 2018) and methods to transform DNA meta-barcoding data to concentrations are less widely used and require important assumptions. We thus performed parallel analyses of phytoplankton composition and concentrations. We use ratios of "diagnostic pigments" to TChla as pigment-based estimates of phytoplankton composition as they are well correlated with other diagnostic pigment approaches for estimating phytoplankton composition and require minimal assumptions (Vidussi et al. 2001; Supporting Information Fig. S5). The seven diagnostic pigments include fucoxanthin (Fuco; biomarker for diatoms), peridinin (Perid; dinoflagellates), monovinyl chlorophyll *b* (MVChlb; chlorophytes), 19′-hexanoyloxyfucoxanthin (Hexfuco; prymnesiophytes), alloxanthin (Allo; cryptophytes) 19′-butanoyloxyfucoxanthin (Butfuco; pelagophytes), and zeaxanthin (Zea; cyanobacteria) (Vidussi et al. 2001). DNA meta-barcoding data were transformed to concentrations by multiplying the relative abundances of all protistan ASVs by concurrently determined POC concentrations, providing estimates of the POC associated with each protistan ASV in each sample (assumptions are detailed in the Discussion). Estimates of POC associated with individual taxonomic groups are not impacted by the phytoplankton identification procedure described above (except forDinophyceae, Cryptophyta, and Haptophyta ASVs; Supporting Information Text S3), though changes to the phytoplankton ASV identification procedure will change compositional estimates. Considering both phytoplankton composition and concentration estimates thus provides two quasi-independent analyses that will be shown to corroborate one another.

Statistical analyses

Known sources of uncertainty in HPLC pigment assessments of phytoplankton composition and concentration include physiological and inter- and intra-lineage variability in biomarker pigmentation (Higgins et al. 2011). To evaluate

the roles of these sources of error in disagreements in phytoplankton composition and concentration estimates, we performed multiple linear regression analyses on each phytoplankton type's composition and concentration residuals relative to the line of best fit determined by Model II linear regression. All predictor and response variables were z-scored prior to regression analyses so that the magnitudes of regression coefficients can be compared to assess their relative importance.

We performed covariation network analysis to determine the patterns of covariation among pigments and both phytoplankton and other protistan classes and ASVs observed via DNA meta-barcoding. 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. Networks were constructed using the Semi-Parametric Rank-based approach for INference in Graphical models (SPRING) method (Yoon et al. 2019). 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), an approach that assumes a sparse network. Our results therefore provide a conservative depiction of the number of associations among pigments and ASVs or classes relative to Pearson correlation-based approaches (Yoon et al. 2019).

We focus on networks constructed considering either class-aggregated or ASV POC concentrations alongside pigment concentrations due to the difficulties associated with robust inference of covariation using compositional data (Yoon et al. 2019). Compositional networks considering either class-aggregated or ASV composition and pigment ratios are included in Supporting Information (Fig. S11) and broadly support the associations observed in concentration networks. Both phytoplankton and other protistan ASVs and classes were included in all networks to determine the phytoplankton assemblages covarying with each pigment, as well as to assess the potential to draw inferences on protistan assemblages more broadly from pigment observations. To maintain consistency with the phytoplankton-specific analyses, phytoplankton and other protists were treated as independent compositions. Where a class included both phytoplankton and other protistan ASVs, the ASVs were separated according to their putative phytoplankton assignments and the prefix "phyto-" is used to distinguish phytoplanktonic ASVs in these classes. Only classes or ASVs found at > 1% relative abundance in at least one sample in their respective compositions were considered in network analyses. Compositions of phytoplankton pigments, phytoplankton ASVs and other protistan ASVs were independently transformed (following Tipton et al. 2018) using the modified centered log-ratio transformation (Yoon et al. 2019) prior to network construction. All analyses presented here consider networks with negative edges (edges indicate significant associations determined by the SPRING method) removed. A community detection algorithm that identifies communities by maximizing within-community interactions and minimizing inter-community interactions (Clauset et al. 2004) was applied to the networks. Clauset et al. (2004) suggest modularity scores > 0.3 indicate "significant" community structure is resolvable from a network.

Results

Dominant phytoplankton taxa in the SBC

Both HPLC and DNA meta-barcoding analysis indicated that the dominant eukaryotic phytoplankton in the SBC are diatoms (Bacillariophyceae), dinoflagellates (Dinophyceae), prymnesiophytes (Prymnesiophyceae), and chlorophytes (Chlorophyta) (Fig. 1). The highest median pigment concentrations and ratios to TChla were observed for Fuco (diatom biomarker), Hexfuco (prymnesiophytes), and MVChlb (chlorophytes), respectively. The dinoflagellate biomarker pigment, Perid, had the 4th highest median ratio to TChla, and the 5th highest median concentration behind the putative cyanobacteria biomarker pigment, Zea. In the DNA meta-barcoding data, dinoflagellates had the highest median composition and concentration, followed closely by diatoms, and then prymnesiophytes and Mamiellophyceae (a class of chlorophytes). Although all classes of Chlorophyta ASVs cumulatively had higher median composition and concentration (7.89% and 9.38 μg L^{-1}) than Prymnesiophyceae ASVs (4.30% and 5.75 μg L^{-1}), the prymnesiophytes comprised a larger proportion of the community than any single class within Chlorophyta. Substantial disagreement across the two methods was found in comparisons of the concentrations and compositions estimated for each of the dominant phytoplankton taxa (Fig. 2). Although diatoms and chlorophytes showed relatively strong correlations ($r^2 > 0.3$, $p < 0.001$) for composition and concentration determinations across the two methods, weaker correlations were observed for prymnesiophytes and dinoflagellates ($r^2 \leq 0.26$; Fig. 2).

Predictor selection to assess sources of disagreement across methods

To assess the contribution of physiological variability in pigmentation to disagreements in phytoplankton composition and concentration estimates across the two methods (Fig. 2), we considered mixed layer depth 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; Behrenfeld et al. 2005). Satellite observations of photosynthetically available radiation were also considered as a covariate of phytoplankton photophysiological variability, but were not significant in predicting most phytoplankton types' residuals. Photosynthetically available radiation and mixed layer depth were significantly, though weakly correlated ($r = -0.32$, $p < 0.001$), but in most regression models, this weak

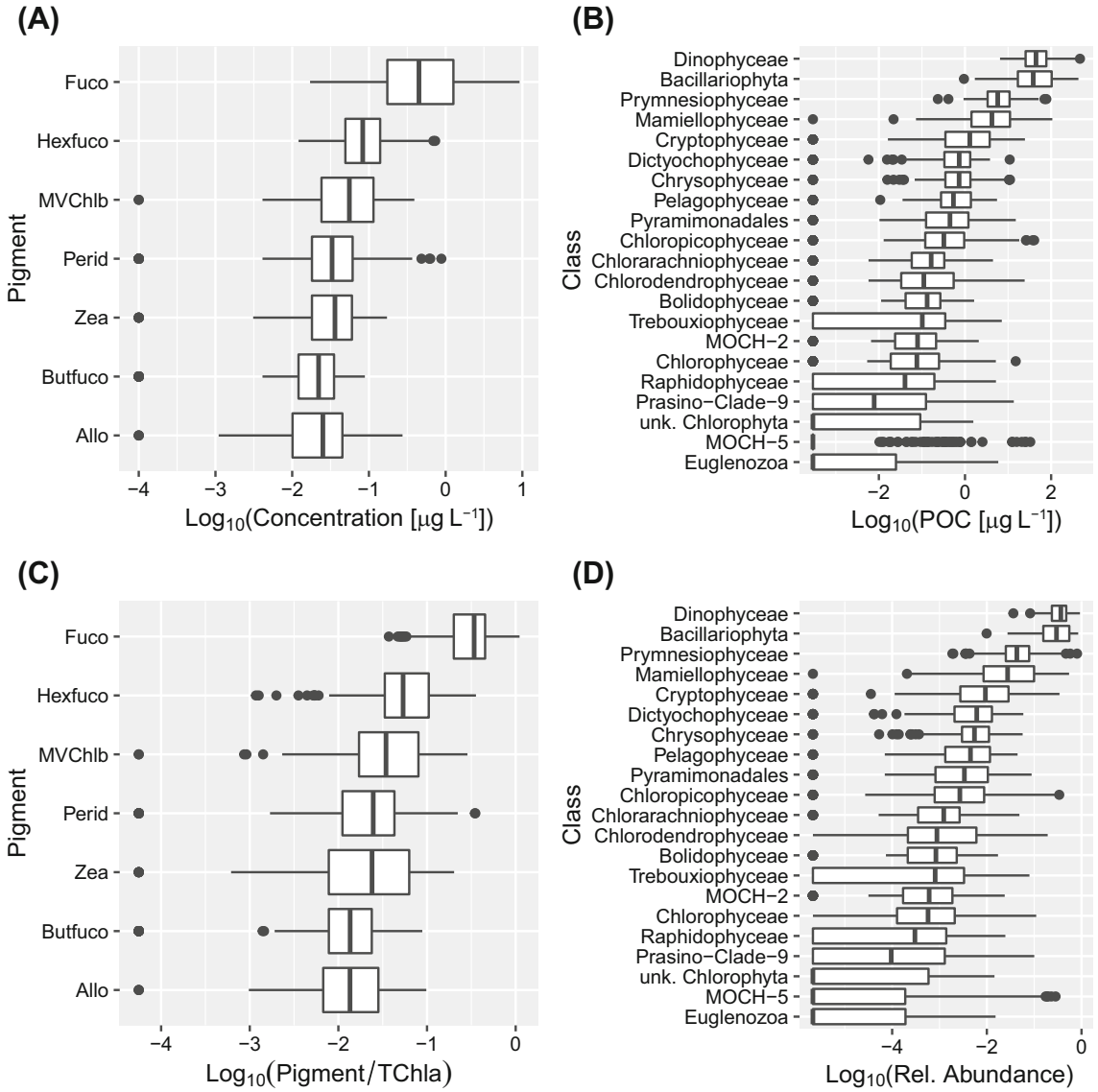

Fig. 1. Distributions of (**A**) phytoplankton biomarker pigment concentrations, (**B**) phytoplankton class POC concentrations, (**C**) biomarker pigment ratios to TChla, and (**D**) phytoplankton class relative abundances observed by (**A**,**C**) HPLC pigment and (**B**,**D**) DNA meta-barcoding analysis.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

collinearity did not substantially impact the values of regression coefficients (Table 1; Supporting Information 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 (Supporting Information Figs. S6–S9). Several Dinophyceae species express the diatom biomarker pigment, Fuco, rather than the typical dinoflagellate biomarker pigment Perid (Zapata et al. 2012). Diatom composition and concentration residuals were correlated with the cumulative composition and concentration of Dinophyceae ASVs ($r = -0.58$ and -0.31, respectively; $p < 0.001$), with most of the variability explained by three Dinophyceae ASVs with compositions that were correlated ($r < -0.4$, $p < 0.001$ for each ASV) with diatom composition residuals. BLASTN searches of these putative Fuco-containing dinoflagellate ASV sequences against the NCBI nucleotide 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; Supporting Information Table S1). The compositions and concentrations of the putative Fuco-containing dinoflagellate ASVs were significantly correlated with Dinophyceae composition and

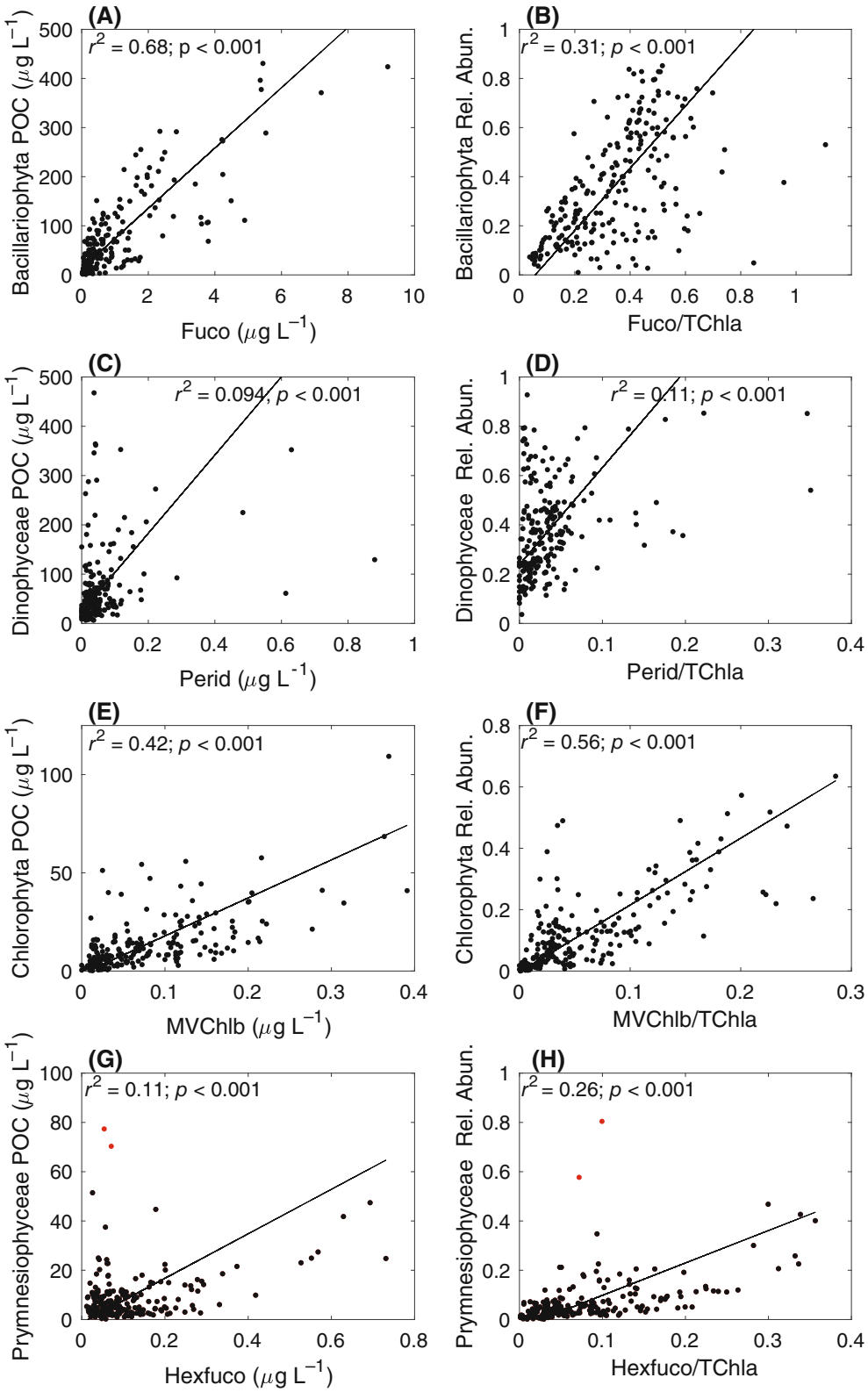

Fig. 2. Comparisons of HPLC pigment and DNA meta-barcoding (**A**,**C**,**E**,**G**) concentration and (**B**,**D**,**F**,**H**) composition estimates for (**A**,**B**) diatoms, (**C**,**D**) dinoflagellates, (**E**,**F**) chlorophytes, and (**G**,**H**) prymnesiophytes. Values of squared Pearson correlation coefficients for each relationship are included in each panel. Lines of best fit are shown in black and were determined by Model II (reduced major axis) regression. In (**G**,**H**), red points show two outlier observations (*see* the main text). Rel. Abun., relative abundance.

Table 1. Multiple linear regression analysis of Model II regression residuals (*see* Fig. 2) for concentrations and compositions of the four dominant Santa Barbara Channel phytoplankton taxa. All variables were z-scored prior to multiple regression analysis. Coefficients of determination (R^2) and sample size (n) are shown for each model. Standardized regression coefficients (p-values) are shown for each predictor included. Fuco-dino is the summed concentration or composition of three putative Fuco-containing dinoflagellate ASVs (*see* the main text; Supporting Information Table S1); Perid-dino is the summed concentration or composition of three putative Perid-containing dinoflagellate ASVs (*see* the main text, Supporting Information Table S1); Chloro-eco1, Chloro-eco2, and Chloro-eco3 is the summed concentration or composition of all ASVs included in Chlorophyta ecotypes 1, 2, or 3, respectively (*see* the main text, Supporting Information Figs. S7, S8); sv15-*P. globosa* is the concentration or composition of the putative *Phaeocystis globosa* ASV associated with underestimation of prymnesiophytes with HPLC pigment methods (*see* the main text, Supporting Information 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	Conc	0.34	183	−0.23 (< 0.001)		−0.51 (< 0.001)					
	Comp	0.36	189			−0.60 (< 0.001)					
Dino	Conc	0.75	183	−0.16 (0.001)		0.72 (< 0.001)	−0.39 (< 0.001)				
	Comp	0.56	189			0.65 (< 0.001)	−0.28 (< 0.001)				
Chloro	Conc	0.55	183	−0.16 (0.002)				0.60 (< 0.001)	0.34 (< 0.001)	−0.39 (< 0.001)	
	Comp	0.52	189					0.69 (< 0.001)	0.20 (< 0.001)	−0.20 (< 0.001)	
Prym	Conc	0.62	183	−0.12 (0.02)	0.28 (< 0.001)						0.70 (< 0.001)
	Comp	0.72	189		0.39 (< 0.001)						0.73 (< 0.001)

Chloro, chlorophyte; Conc, concentration; Comp, composition; Dino, dinoflagellate; MLD, mixed layer depth; PAR, photosynthetically available radiation; Prym, prymnesiophyte.
*Concentrations were used to predict concentration residuals, while compositions were used to predict composition residuals.

concentration residuals (r = 0.64 and 0.69, respectively; $p < 0.001$). Similarly, two putative Perid-containing Dinophyceae ASVs were identified as those with compositions most strongly correlated ($r < -0.4$, $p < 0.001$) with Dinophyceae composition residuals. Again, BLASTN searches of these ASV sequences against the NCBI nucleotide database suggested they were derived from a Perid-containing genus (*Tripos*; Supporting Information Table S1).

Investigations of the potential for intra-lineage variability in biomarker pigmentation to contribute to error in pigment-based taxon concentrations and compositions showed that Chlorophyta composition and concentration residuals varied systematically with the dominant Chlorophyta class within each sample (Supporting Information Fig. S7). We grouped chlorophyte classes into three "ecotypes" based on these systematic differences. 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 (associated with residuals with a distribution centered near 0), and ecotype 3 included Pyramimonadales (associated with negative residual values). Further analysis showed the concentration and composition of ecotype 1 exhibited a strong linear relationship with Chlorophyta concentration and composition residuals, while the concentrations and compositions of ecotypes 2 and 3 showed weaker systematic variations with Chlorophyta concentration and composition residuals (Supporting Information Fig. S8).

Systematic variation of prymnesiophyte concentration and composition residuals with the dominant Prymnesiophyceae order was not observed (Supporting Information Fig. S7). However, two outliers were identified in comparisons of Prymnesiophyceae concentrations and compositions across the two methods (shown in red in Fig. 2G,H). These outliers originated from a single cruise in April 2011, and were dominated by an ASV assigned as *Phaeocystis globosa* that was highly correlated with prymnesiophyte composition and concentration residuals (Supporting Information Fig. S9; r = 0.72 and 0.73 for compositions and concentrations, respectively; $p < 0.001$). BLASTN searches against the NCBI nucleotide database showed this ASV sequence was a perfect match to 18S rDNA sequences from several strains of *P. globosa*, corroborating the ensemble taxonomic assignment.

Multiple linear regression analysis to characterize sources of disagreement across methods

Table 1 shows multiple linear regression statistics determined for predictions of concentration and composition residuals based on linear combinations of the presumed physiological correlates (mixed layer depth and photosynthetically available radiation) and the various sources of inter- and intra-lineage variability in biomarker pigmentation identified above. Additional multiple linear regression results are included in Supporting Information Table S2 to support the

conclusions drawn here. Mixed layer depth was a significant predictor of the concentration residuals of all four dominant phytoplankton taxa (Table 1). Negative mixed layer depth coefficient values in all concentration residual regression models indicated that a deeper mixed layer was associated with higher pigment : POC ratios as expected. Mixed layer depth was insignificant in predicting the composition residuals of diatoms and dinoflagellates but significant in predicting the composition residuals of chlorophytes and prymnesiophytes, though R^2 values only decreased by 0.02 when mixed layer depth was excluded from these models (Supporting Information Table S2). Photosynthetically available radiation was a significant predictor of prymnesiophyte concentration and composition residuals, with positive regression coefficient values indicating a decrease in the Hexfuco to Prymnesiophyceae POC ratio with increasing irradiance. Photosynthetically available radiation was also a significant predictor of diatom composition residuals, though removing photosynthetically available radiation from these regression models again resulted in a minor (3%) decrease in R^2 values (Supporting Information Table S2). The reduced importance of physiological correlates in predicting phytoplankton composition residuals implies similarities in variations in the expression of TChla and biomarker pigments in response to environmental stimuli across most of the dominant phytoplankton taxa.

Sources of inter-lineage (for diatoms) and intra-lineage (for all other taxa) variability in biomarker pigmentation had standardized regression coefficient values that were more than twofold higher in magnitude than the presumed physiological correlates (mixed layer depth and photosynthetically available radiation) in all regression models except the prymnesiophyte composition residual model. As expected, R^2 values were generally lower when predicting residuals where a taxon's pigment-based concentrations and compositions were better-correlated with DNA meta-barcoding concentrations and compositions (Table 1; Fig. 2). Regression models that predicted diatom concentration residuals using mixed layer depth and the summed concentration of three putative Fuco-containing dinoflagellate ASVs (Supporting Information Table S1) yielded R^2 values of 0.34 (Table 1). An R^2 value of 0.36 was found when predicting diatom composition residuals with only the composition of three putative Fuco-containing dinoflagellate ASVs. For Dinophyceae concentration residuals, R^2 values of 0.75 were obtained in a regression model including mixed layer depth and the concentrations of both putative Fuco- and Perid-containing dinoflagellates (Table 1). Like diatom concentration residuals, putative Fuco- and Perid-containing dinoflagellate concentrations had substantially larger standardized regression coefficient values than mixed layer depth. Models including the composition of putative Fuco- and Perid-containing dinoflagellates also fit dinoflagellate composition residuals well (R^2 = 0.56; Table 1; Supporting Information Fig. S10).

Models that predicted Chlorophyta concentration and composition residuals from the cumulative concentration and composition of each of the three Chlorophyta ecotypes achieved R^2 values of 0.55 and 0.52, respectively (models used for concentration residuals also included mixed layer depth; Table 1). Furthermore, the composition and concentration of each Chlorophyta ecotype was significant in predicting Chlorophyta composition and concentration residuals, with coefficients following the trends expected based on Supporting Information Figs. S7, S8. Chlorophyta ecotype 1 was associated with regression coefficient values with the highest magnitude in both models. Models including the two physiological correlates along with the concentration or composition of the putative *P. globosa* ASV predicted prymnesiophyte concentration and composition residuals well (R^2 = 0.72 and 0.62, respectively).

Network analysis to characterize covariation of phytoplankton classes and ASVs with biomarker pigment concentrations

We first consider phytoplankton subnetworks (a subset of the network with non-phytoplanktonic members removed) including only the seven diagnostic pigments and phytoplankton class or ASV concentrations (Fig. 3; Supporting Information Fig. S11) and expand our discussion to consider networks including both phytoplankton and other protistan classes and ASVs below (*see* "Discussion" section; Fig. 4; Supporting Information Figs. S14–S16). 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 linked either directly by positive edges shared with one another, or indirectly through positive edges shared with common nodes. A modularity score of 0.44 was found for the phytoplankton pigment and class subnetwork.

Figure 3A shows the phytoplankton class and biomarker pigment concentration subnetwork. In general, biomarker pigment concentrations covaried 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 Hexfuco. Dinophyceae shared positive edges with both Fuco and Perid, but only shared community membership with Fuco. Interestingly, Zea, rather than the typical chlorophyte biomarker MVChlb, shared a positive edge with the Chlorophyta classes Trebouxiophyceae and Chloropicophyceae, and was found in the same community as Chlorodendrophyceae and

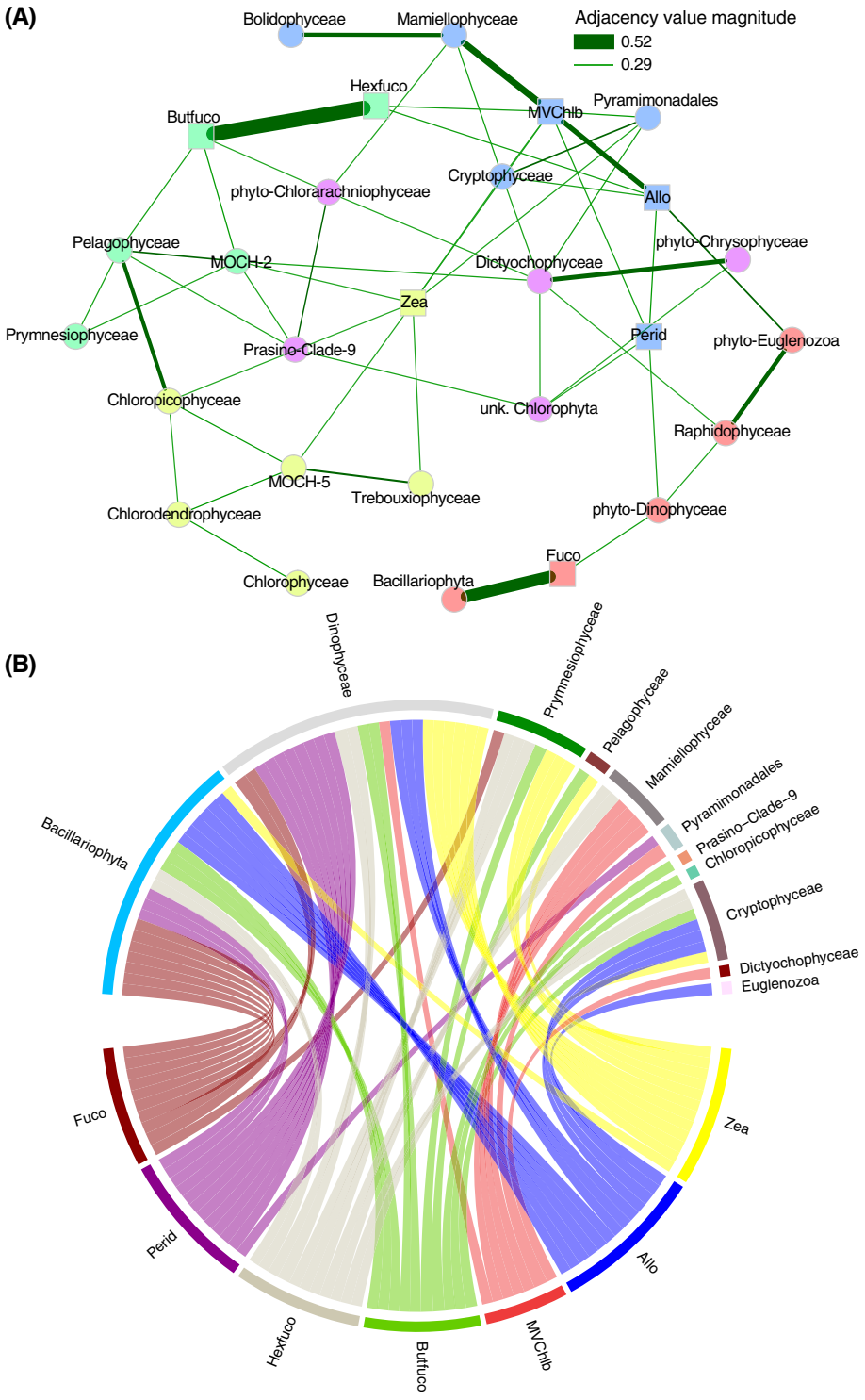

Fig. 3. Associations among phytoplankton biomarker pigment concentrations and phytoplanktonic classes and ASVs determined from the covariation network including both phytoplankton and other protists. (**A**) The phytoplankton class (circles) and biomarker pigment (squares) concentration subnetwork. Nodes 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 ASVs concentrations. MOCH, marine ochrophytes; unk., unknown.

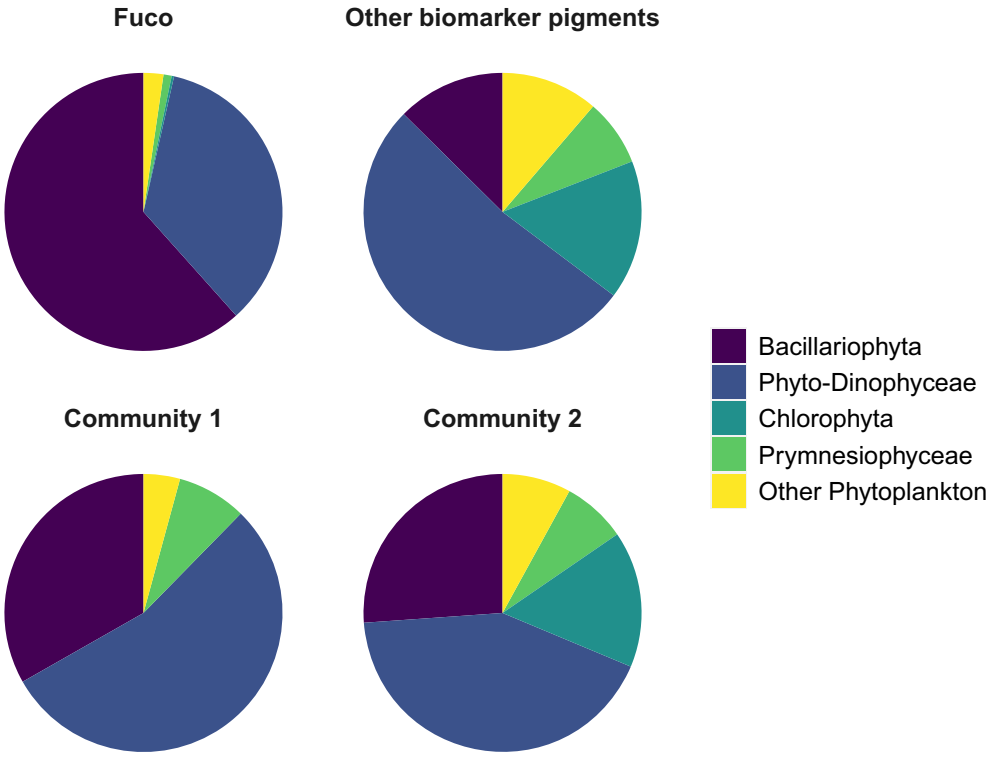

Fig. 4. Mean composition of phytoplankton ASVs found within the four communities identified by the community detection algorithm (Clauset et al. 2004). Communities 1 and 2 do not include any of the seven pigments considered in network analysis. "Other Phytoplankton" includes ASVs representative of Euglenozoa, Chrysophyceae, Dictyophyceae, Cryptophyceae, Raphidophyceae, Chlorarachniophyceae, Pelagophyceae, and MOCH (a phylogenetically distinct lineage of marine ochrophytes). Fuco, fucoxanthin; MOCH, marine ochrophytes.

Chlorophyceae. These classes comprised Chlorophyta ecotype 1 defined above, which was associated with consistent underestimation of Chlorophyta concentrations and compositions from HPLC pigment data (Table 1; Supporting Information Figs. S7, S8).

Five communities of phytoplankton pigments and classes were identified from the phytoplankton subnetwork (Fig. 3A), four of which included at least one pigment. Associations among the pigments 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. Strong covariation and shared community membership were found between the chlorophyte and cryptophyte biomarker pigments (MVChlb and Allo) and the prymnesiophyte and pelagophyte biomarker pigments (Hexfuco and Butfuco; Fig. 3A). Unexpectedly, Perid was found in the same community as MVChlb and Allo. Four phytoplankton classes neither shared an edge nor community membership with a pigment.

Figure 3B shows a chord diagram (Gu et al. 2014) illustrating the direct associations between phytoplankton pigments and ASVs. A diverse assemblage of phytoplankton ASVs covaried with each pigment. The largest proportion of ASVs associated with each pigment often belonged to the phytoplankton class for which the pigment is the corresponding biomarker. For example, Fuco shared positive edges with seven Bacillariophyta ASVs, two Dinophyceae ASVs, and one Prymnesiophyceae ASV, while Perid shared positive edges with eight Dinophyceae ASVs, three Bacillariophyta ASVs, and one Pyramimonadales ASV. Similarly, MVChlb shared positive edges with five Chlorophyta ASVs (four of class Mamiellophyceae and one of Pyramimonadales), and two ASVs representative of Dinophyceae and Dictyochophyceae. Unexpectedly, Butfuco and Hexfuco shared positive edges with 1–3 ASVs from a diverse array of phytoplankton classes. Nearly half (6 of 13) of the positive edges between Allo and phytoplankton ASVs were associated with Bacillariophyceae ASVs rather than Cryptophyceae ASVs.

Four communities were identified from the pigment and ASV concentration network (Fig. 4). One community included Fuco, the other six pigments were found in a 2nd community, and two communities did not include a pigment (henceforth Communities 1 and 2). In total, 176 of 313 total phytoplankton ASVs included in network analyses shared community membership with a biomarker pigment, with 78 ASVs found in the Fuco community and 98 ASVs found in the community with all other pigments. As expected, the Fuco community was dominated by Bacillariophyta ASVs and had the lowest mean composition of Prymnesiophyceae, Chlorophyta, and diverse other phytoplankton ASVs across the four communities. Mean composition of the community including the other

six pigments was dominated by Dinophyceae ASVs, had relatively high contributions of Prymnesiophyceae, Chlorophyta, and diverse other phytoplankton ASVs, and had the lowest diatom contributions of any of the four communities. One of the communities that lacked a pigment (Community 2) had similar mean composition to the community including six pigments, while the other (Community 1) had an unusual composition that included no Chlorophyta ASVs.

Discussion

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

The application of HPLC pigment observations to quantify phytoplankton composition and concentrations assumes that physiological and inter- and intra-lineage sources of variability in biomarker pigmentation are negligible. However, our results show that this is not the case. We identified inter- and intra-lineage variability in biomarker pigmentation as the largest source of uncertainty in pigment estimates of the four dominant phytoplankton taxa in the SBC (Fig. 2; Table 1; Supporting Information Figs. S6–S10). Similar observations in other systems suggest this may be a consistent issue in the coastal ocean (Goericke and Montoya 1998; Irigoien et al. 2004). Mixed layer depth and photosynthetically available radiation, used here as proxies for physiological variations in pigment expression, were significant, though less important, predictors of disagreement between pigment and DNA meta-barcoding estimates of concentrations and composition for this highly productive, coastal site (Fig. 2; Table 1).

Many assumptions are required to estimate phytoplankton concentration and composition with DNA meta-barcoding. DNA meta-barcoding is assumed to provide precise and accurate estimates of phytoplankton composition in all studies where these data are treated quantitatively, though evaluations of uncertainty in DNA meta-barcoding workflows are difficult and rarely performed. Recent attempts to validate these methods suggest that DNA meta-barcoding achieves precision comparable to HPLC pigment analysis (Catlett et al. 2020a; Yeh et al. 2021). Our finding that inter- and intra-lineage variability in pigmentation is the dominant source of disagreement in pigment and DNA meta-barcoding estimates of phytoplankton composition (Figs. 1, 2; Table 1; Supporting Information Figs. S6–S10) further supports the use of DNA meta-barcoding as a method with comparable (or greater) accuracy and precision to other accepted methods.

A major limitation of DNA meta-barcoding is the compositionality (relative abundance) constraint (Gloor et al. 2017; Lin et al. 2019). To integrate DNA meta-barcoding with other methods to quantify phytoplankton composition, this constraint must be addressed either with robust identification of phytoplankton ASVs, or with a transformation of ASV compositions to concentrations. Both were performed here and require assumptions. Overall, our analysis shows that phytoplankton ASV identification is complicated by the prevalence of mixotrophy in marine protists (Mitra et al. 2016) combined with limited phylogenetic resolution of short meta-barcode amplicons and inadequate taxonomy prediction methods (Supporting Information Table S1; Figs. S1, S2). Furthermore, many of the putative trophic mode assignments compiled here rely on the assumption that taxonomic groups share common phenotypes (Adl et al. 2019). One example where this assumption is not satisfied is among the Dinophyceae (Adl et al. 2019). Here, limited ability to assign Dinophyceae ASVs to trophic groups necessitated classifying unknown Dinophyceae ASVs as phytoplankton; however, our ability to explain most of the disagreement in Dinophyceae concentration and composition estimates across the two methods demonstrates that this assumption is valid for this data set (Fig. 2; Supporting Information Table S1; Figs. S3, S4). The abundance of unknown Dinophyceae ASVs in our data also highlights the limited resolution of short amplicons and standard taxonomic assignments. This problem persisted after implementing ensemble methods (Catlett et al. 2021*b*) that increase the specificity of ASV taxonomic assignments (likely at the expense of increased false-positive annotations; Supporting Information Fig. S1; Murali et al. 2018; Catlett et al. 2021*b*). For example, one of the putative Fuco-containing dinoflagellate ASVs identified here (sv8) was classified as unknown Dinophyceae and was a perfect match to representatives from several dinoflagellate genera, only some of which express Fuco (Supporting Information Table S1). This problem is partially attributable to the limited taxonomic resolution of short amplicons like the 18S-V9 amplicon considered here, though some studies suggest analysis of longer amplicons results in less accurate composition estimates (Bradley et al. 2016).

To circumvent the compositionality constraint in DNA meta-barcoding data, we scaled protistan relative abundances to concurrent observations of POC concentrations to estimate ASV and group concentrations. This approach relies on two key assumptions: that variability in POC concentrations is directly proportional to variability in protist community biomass and that variability in protistan 18S rDNA copy number scales with cell biomass. The 1st assumption is specific to our analyses that focus on characterizing variability in (rather than the magnitude of) taxon POC concentrations. In the SBC, relatively strong covariation between monovinyl Chl *a* and POC concentrations, in conjunction with an absence of covariation of cyanobacterial pigments with POC (Supporting Information Fig. S12), suggest that cyanobacterial contributions to POC variability are negligible. Free-living bacterioplankton biomass contributions to total POC are also negligible (mean = 4.8%, standard deviation = 2.6%) in this data set (Supporting Information Text S4; Table S3). However, the significance of detrital and protist-associated prokaryotic contributions to POC variability, as well as variability in the

proportion of DNA-containing detritus of protistan origin, remains uncertain in most marine systems. Differences in filter pore sizes used here for protistan DNA meta-barcoding (1.2 μm) and POC analysis (0.7 μm nominal pore size) introduce further complications. The 2nd assumption is supported by the well-documented 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 Lessard 2000), across diverse marine protist lineages. The agreement observed with HPLC pigment methods after accounting for known sources of error in pigment composition and concentration estimates indicates that the POC-scaling of protistan relative abundances is appropriate for the analyses performed here, though remaining disagreements across the methods may be attributed to error introduced by this scaling. Further study should be devoted to validating this approach against internal standard methods (Lin et al. 2019). Overall, integrating HPLC pigment and DNA meta-barcoding analysis allowed us to identify weaknesses in HPLC pigment phytoplankton composition estimates and provided partial validation for our treatment of DNA meta-barcoding data.

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 pigment expression and variability in physiological status are significant sources of error in HPLC estimates of phytoplankton concentration and composition in the SBC (Fig. 2; Table 1; Supporting Information Figs. S6–S10). Intra-lineage variability in pigment expression in smaller sized phytoplankton including prymnesiophytes and chlorophytes has not been previously observed in studies that integrate HPLC and microscopic observations (Irigoien et al. 2004). Here, we contextualize our observations of physiological and inter- and intra-lineage variability in pigmentation, and in turn, demonstrate that integrating pigment and DNA meta-barcoding observations provides novel insights into phytoplankton physiology and ecology.

Inter-lineage variability in biomarker pigmentation was the primary source of error in diatom concentration estimates from pigments, while intra-lineage variability led to uncertainty in estimated concentrations and compositions of dinoflagellates, chlorophytes, and prymnesiophytes (Fig. 2; Table 1). Our results suggest that in the SBC, Fuco-containing dinoflagellates (Supporting Information Table S1) often lead to an over-estimation of diatom concentrations and compositions and an underestimation of dinoflagellate concentrations and compositions with pigment methods (Fig. 2; Table 1; Supporting Information Fig. S6). Some of the putative Fuco-containing genera identified here (e.g., *Karenia*; Supporting Information Table S1) have not been observed previously in the SBC, although observations of adjacent waters have detected and noted high concentrations of some *Gymnodinium* species (Cullen et al. 1982). In network analyses (Figs. 3, 4), Fuco- and Perid-containing dinoflagellate ASVs (Supporting Information 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).

Intra-lineage variability in biomarker pigmentation was the dominant source of uncertainty in HPLC pigment estimates of chlorophyte and prymnesiophyte concentrations and compositions. Dominance by one of four classes within Chlorophyta (Trebouxiophyceae, Chlorodendrophyceae, Chlorophyceae, Chloropicophyceae, comprising ecotype 1) was associated with consistent underestimation of Chlorophyta concentrations and compositions with pigment methods (Fig. 2; Table 1; Supporting Information Figs. S7, S8). Although 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, 2017). Interestingly, inspection of the concentration and composition distributions of each ecotype according to the month sampled (Supporting Information Fig. S13) showed that ecotype 1 tended to exhibit the highest composition and concentrations during the summer (July and August), a time of year that is likely associated with a reduction in pigment : carbon ratios due to high surface irradiance and stratification and low surface nutrient concentrations in the SBC (Catlett et al. 2021*a*). Network analysis also confirmed that both Mamiellophyceae and Pyramimonadales (Chlorophyta ecotypes 2 and 3) covary with MVChlb while Chlorophyta classes in ecotype 1 covary more strongly with the photoprotective pigment, Zea (Fig. 3A). Thus, the intra-Chlorophyta variability in MVChlb expression is likely due to a combination of genetic and physiological variability in pigment : biomass ratios.

A single putative *P. globosa* ASV (sv15) consistently led to an underestimation of Prymnesiophyceae concentrations and compositions with pigments (Table 1; Supporting Information Fig. S10). Notably, this ASV was directly associated with Fuco in the network analysis (Fig. 3B). *P. globosa* forms large carbon-rich colonies and, when exposed to high irradiance, expresses little to no Hexfuco but continues to express Fuco (Schoemann et al. 2005). These observations are consistent with the dependence of Prymnesiophyceae residuals on photosynthetically available radiation (Fig. 2; Table 1) and previous microscopy observations of *Phaeocystis* colonies in the SBC (Goodman et al. 2012). Taken together, this suggests *P. globosa* is an important colony-forming prymnesiophyte in the SBC that is unaccounted for with pigment methods.

In addition to inter- and intra-lineage variability in pigmentation, mixed layer depth was a significant predictor of the concentration residuals of all four of the dominant SBC phytoplankton types (Table 1). The negative coefficient values for mixed layer depth in all concentration residual regression analyses indicate deeper mixed layers, which are typically associated with reduced irradiance and increased nutrient availability, were associated with increased pigment : carbon ratios. Interestingly, mixed layer depth was an insignificant predictor of diatom and dinoflagellate composition residuals, and was only associated with marginal increases (< 0.025) in R^2 values in prymnesiophyte and chlorophyte composition residual regression models (Supporting Information Table S2). Photosynthetically available radiation was a significant predictor of prymnesiophyte composition and concentration residuals but was insignificant in predicting most of the other residuals considered here (Supporting Information Table S2), possibly due to heightened sensitivity of *P. globosa* Hexfuco expression to irradiance (Schoemann et al. 2005). Overall, these correlates of phytoplankton physiological status were more important predictors of concentration rather than composition residuals (Table 1; Supporting Information Table S2). This implies that physiological responses to environmental stimuli are a less important source of uncertainty in pigment-based phytoplankton compositions relative to concentrations as many taxa apparently alter their expression of TChla and accessory pigments in similar ways. Altogether, these results show that integration of DNA meta-barcoding and HPLC pigment data provides novel insights into the physiological responses of dominant phytoplankton types to environmental stimuli.

Implications for satellite remote sensing of plankton assemblages

Perhaps the greatest strength in HPLC pigment assessments of phytoplankton composition 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 pigment concentrations and/or derived composition indices from remotely sensible bio-optical properties (Uitz et al. 2015; Chase et al. 2017). The forthcoming launch of the NASA Plankton, Aerosols, Clouds, ocean Ecosystems mission (Werdell et al. 2019) is expected to improve retrievals of pigment concentrations from satellite ocean color observations by improving resolution of small-scale phytoplankton absorption features (Uitz et al. 2015; Catlett and Siegel 2018). Our results thus have important implications for future efforts to observe and interpret phytoplankton composition via remotely sensible phytoplankton pigment concentrations.

Despite uncertainty in pigment-based phytoplankton composition and concentration estimates (Fig. 2; Table 1), the network analysis applied here revealed that diverse communities of phytoplankton classes and ASVs are associated with different biomarker pigments (Figs. 3, 4; Supporting Information Fig. S11). This suggests that characterizing the assemblages of phytoplankton that covary with distinct suites of biomarker pigments may offer an alternative path to observing phytoplankton composition from ocean color, circumventing the assumptions linking a pigment to compositions or concentrations of a phytoplankton taxon. Although further research is needed to evaluate whether the patterns observed here can be extrapolated to other regions of the world's oceans, improved understanding of the ecological processes that drive this covariability may provide a foundation for future development of predictive models that estimate phytoplankton composition and concentrations from pigments and other remotely sensible oceanographic properties. Furthermore, recent work documenting the numerous potential interactions (inferred from covariability) within and between phytoplankton, prokaryotic, protistan, and metazoan communities (Lima-Mendez et al. 2015; Berdjeb et al. 2018) highlights an additional benefit of this approach: it can be readily extended to characterize the broader planktonic communities that covary with distinct suites of biomarker pigments.

To illustrate this point, we performed a preliminary assessment of the extent to which non-phytoplanktonic protists detected in our DNA meta-barcoding data covaried with biomarker pigments by including non-phytoplanktonic classes and ASVs in network analyses (*see* "Methods" section). Different suites of pigments covaried with diverse and distinct assemblages of putative microzooplankton grazers, parasites, and other protists (Supporting Information Figs. S14–S16). Although associations between pigments and non-phytoplanktonic ASVs and classes were typically weaker than those observed between pigments and phytoplankton, networks that included non-phytoplanktonic ASVs and classes remained fully connected, meaning all pigments and protistan classes or ASVs were either directly or indirectly (via an intermediate node) associated with one another. Overall, this preliminary analysis confirms that characterizing the covariability among pigments and both phytoplankton and protistan assemblages provides a path to link the structure of protistan assemblages and ecosystems to satellite ocean color observations.

However, several major limitations and questions and a great deal of research remain to address this goal. Identifying correlation, covariation, or co-presence is not equivalent to directly verifying and identifying the nature, frequency, and dynamics of associations and interactions among organisms (Fuhrman et al. 2015). To remotely sense plankton assemblages, a predictive understanding of the processes governing the associations and interactions among pigments and protistan classes and ASVs observed here and elsewhere is required (Lima-Mendez et al. 2015; Berdjeb et al. 2018). Application of analyses like those performed here to a broader array of ocean ecosystems and spatiotemporal scales will determine the extent to which the associations among pigments and

protistan classes and ASVs observed here vary. Although the optimal observational scales of time, space, and plankton diversity for predicting plankton composition and concentrations from phytoplankton pigment concentrations remain unknown, community-oriented analyses offer a path to address these questions and may provide a path to improve marine ecosystem monitoring from ocean color remote sensing.

Conflict of interest

The authors declare no conflicts of interest.

Data availability statement

Curated data used in the present study are available via Environmental Data Initiative (Catlett et al. 2020*b*, 2022). Raw sequence data are available in the National Center for Biotechnology Information's Sequence Read Archive (accession number PRJNA532583).

References

Adl, S. M., and others. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. J. Eukaryot. Microbiol. **66**: 4–119. doi:10.1111/jeu.12691

Aitchison, J. 1982. The statistical analysis of compositional data. J. R. Stat. Soc. B. Methodol. **44**: 139–160.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2

Behrenfeld, M. J., E. Boss, D. A. Siegel, and D. M. Shea. 2005. Carbon-based ocean productivity and phytoplankton physiology from space. Global Biogeochem. Cycl. **19**. doi: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. ISME J. **12**: 1907–1917. doi:10.1038/s41396-018-0097-x

Bradley, I. M., A. J. Pinto, and J. S. Guest. 2016. Design and evaluation of Illumina MiSeq-compatible, 18 S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. Appl. Environ. Microbiol. **82**: 5878–5891. doi:10.1128/AEM.01630-16

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. J. Geophys. Res. Oceans **109**. doi:10.1029/2004jc002419

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. Nat. Methods **13**: 581–583. doi:10.1038/nmeth.3869

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. J. Geophys. Res. Oceans **123**: 246–264. doi:10.1002/2017JC013195

Catlett, D., P. G. Matson, C. A. Carlson, E. G. Wilbanks, D. A. Siegel, and M. D. Iglesias-Rodriguez. 2020*a*. Evaluation of accuracy and precision in an amplicon sequencing workflow for marine protist communities. Limnol. Oceanogr. Methods **18**: 20–40. doi:10.1002/lom3.10343

Catlett, D., D. A. Siegel, and N. Guillocheau. 2020*b*. Plumes and Blooms: Curated oceanographic and phytoplankton pigment observations ver 1. Environ. Data Initiat. doi:10.6073/pasta/f88ee1dc32b8785fe6ce57d80722e78c

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

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

Catlett, D., D. A. Siegel, P. G. Matson, E. K. Wear, C. A. Carlson, T. S. Lankiewicz, and M. D. Iglesias-Rodriguez. 2022. Plumes and Blooms: Microbial eukaryote diversity and composition ver 1. Environ. Data Initiat. doi:10.6073/pasta/bbe75db57c4ce7f7842e22923ae6badf

Chase, A., E. Boss, I. Cetinić, and W. Slade. 2017. Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra: Toward a global algorithm. J. Geophys. Res. Oceans **122**: 9725–9743. doi:10.1002/2017JC012859

Clauset, A., M. E. Newman, and C. Moore. 2004. Finding community structure in very large networks. Phys. Rev. E **70**: 066111. doi:10.1103/PhysRevE.70.066111

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. doi:10.1128/AEM.72.4.2496-2506.2006

Csardi, G., and T. Nepusz. 2006. The igraph software package for complex network research. Int. J. Comp. Syst. **1695**: 1–9.

Cullen, J. J., S. G. Horrigan, M. E. Huntley, and F. M. Reid. 1982. Yellow water in La Jolla Bay, California, July 1980. I. A bloom of the dinoflagellate, *Gymnodinium flavum* Kofoid & Swezy. J. Exp. Mar. Biol. Ecol. **63**: 67–80. doi:10.1016/0022-0981(82)90051-X

De Vargas, C., and others. 2015. Eukaryotic plankton diversity in the sunlit ocean. Science **348**: 1261605. doi:10.1126/science.1261605

Fuhrman, J. A., J. A. Cram, and D. M. Needham. 2015. Marine microbial community dynamics and their ecological interpretation. Nat. Rev. Microbiol. **13**: 133–146. doi:10.1038/nrmicro3417

Gloor, G. B., J. M. Macklaim, V. Pawlowsky-Glahn, and J. J. Egozcue. 2017. Microbiome datasets are compositional:

And this is not optional. Front. Microbiol. **8**: 2224. doi:10.3389/fmicb.2017.02224

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. doi:10.1128/AEM.01298-08

Goericke, R., and J. P. Montoya. 1998. Estimating the contribution of microalgal taxa to chlorophyll a in the field—Variations of pigment ratios under nutrient-and light-limited growth. Mar. Ecol. Prog. Ser. **169**: 97–112. doi:10.3354/meps169097

Gong, W., N. Hall, H. Paerl, and A. Marchetti. 2020. Phytoplankton composition in a eutrophic estuary: Comparison of multiple taxonomic approaches and influence of environmental factors. Environ. Microbiol. **22**: 4718–4731. doi: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. Cont. Shelf Res. **48**: 27–39. doi:10.1016/j.csr.2012.08.011

Gu, Z., L. Gu, R. Eils, M. Schlesner, and B. Brors. 2014. Circlize implements and enhances circular visualization in R. Bioinformatics **30**: 2811–2812. doi:10.1093/bioinformatics/btu393

Guidi, L., and others. 2016. Plankton networks driving carbon export in the oligotrophic ocean. Nature **532**: 465–470. doi:10.1038/nature16942

Guillou, L., and others. 2012. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. **41**: D597–D604. doi:10.1093/nar/gks1160

Higgins, H. W., S. W. Wright, and L. Schluter. 2011. Quantitative interpretation of chemotaxonomic pigment data, p. 257–313. *In* S. Roy, C. A. Llewellyn, E. S. Egeland, and G. Johnsen [eds.], Phytoplankton pigments: Characterization, chemotaxonomy and applications in oceanography.

Huson, D. H., A. F. Auch, J. Qi, and S. C. Schuster. 2007. MEGAN analysis of metagenomic data. Genome Res. **17**: 377–386. doi:10.1101/gr.5969107

Irigoien, X., B. Meyer, R. Harris, and D. Harbour. 2004. Using HPLC pigment analysis to investigate phytoplankton taxonomy: The importance of knowing your species. Helgoland Mar. Res. **58**: 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, p. 3–77. *In* S. Roy, C. A. Llewellyn, E. S. Egeland, and G. Johnsen [eds.], Phytoplankton pigments: Characterization, chemotaxonomy and applications in oceanography.

Kramer, S. J., and D. A. Siegel. 2019. How can phytoplankton pigments be best used to characterize surface ocean phytoplankton groups for ocean color remote sensing algorithms? J. Geophys. Res. Oceans **124**: 7557–7574. doi:10.1029/2019JC015604

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 Res. Part II **45**: 2133–2170. doi:10.1016/s0967-0645(98)00066-6

Lima-Mendez, G., and others. 2015. Determinants of community structure in the global plankton interactome. Science **348**: 1262073. doi:10.1126/science.1262073

Lin, Y., N. Cassar, A. Marchetti, C. Moreno, H. Ducklow, and Z. Li. 2017. Specific eukaryotic plankton are good predictors of net community production in the Western Antarctic Peninsula. Sci. Rep. **7**: 14845. doi:10.1038/s41598-017-14109-1

Lin, Y., S. Gifford, H. Ducklow, O. Schofield, and N. Cassar. 2019. Towards quantitative microbiome community profiling using internal standards. Appl. Environ. Microbiol. **85**: e02634-18. doi:10.1128/AEM.02634-18

Lombard, F., and others. 2019. Globally consistent quantitative observations of planktonic ecosystems. Front. Mar. Sci. **6**: 196. doi:10.3389/fmars.2019.00196

Lopes dos Santos, A., P. Gourvil, F. Rodríguez, J. L. Garrido, and D. Vaulot. 2016. Photosynthetic pigments of oceanic Chlorophyta belonging to prasinophytes clade VII. J. Phycol. **52**: 148–155. doi:10.1111/jpy.12376

Lopes dos Santos, A., and others. 2017. Chloropicophyceae, a new class of picophytoplanktonic prasinophytes. Sci. Rep. **7**: 1–20. doi:10.6084/m9.figshare.5027375

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. Mar. Ecol. Prog. Ser. **144**: 265–283. doi:10.3354/meps144265

Meinshausen, N., and P. Bühlmann. 2006. High-dimensional graphs and variable selection with the lasso. Ann. Stat. **34**: 1436–1462. doi:10.1214/009053606000000281

Menden-Deuer, S., and E. J. Lessard. 2000. Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. Limnol. Oceanogr. **45**: 569–579. doi:10.4319/lo.2000.45.3.0569

Mitra, A., and others. 2016. Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: Incorporation of diverse mixotrophic strategies. Protist **167**: 106–120. doi:10.1016/j.protis.2016.01.003

Murali, A., A. Bhargava, and E. S. Wright. 2018. IDTAXA: A novel approach for accurate taxonomic classification of microbiome sequences. Microbiome **6**: 1–14. doi:10.1186/s40168-018-0521-5

Peschel, S., C. L. Müller, E. von Mutius, A. L. Boulesteix, and M. Depner. 2021. NetCoMi: Network construction and comparison for microbiome data in R. Brief. Bioinform. **22**: bbaa290. doi:10.1093/bib/bbaa290

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 Res. **41**: D590–D596. doi:10.1093/nar/gks1219

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

Schlüter, L., F. Møhlenberg, H. Havskum, and S. Larsen. 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. Mar. Ecol. Prog. Ser. **192**: 49–63. doi:10.3354/meps192049

Schoemann, V., S. Becquevort, J. Stefels, V. Rousseau, and C. Lancelot. 2005. Phaeocystis blooms in the global ocean and their controlling mechanisms: A review. J. Sea Res. **53**: 43–66. doi:10.1016/j.seares.2004.01.008

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 Res. Part II **112**: 117–128. doi:10.1016/j.dsr2.2014.02.006

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

Uitz, J., D. Stramski, R. A. Reynolds, and J. Dubranna. 2015. Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. Remote Sens. Environ. **171**: 58–74. doi:10.1016/j.rse.2015.09.027

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. J. Chromatogr. A **910**: 31–49. doi:10.1016/S0378-4347(00)00603-4

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. J. Geophys. Res. Oceans **106**: 19939–19956. doi:10.1029/1999JC000308

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. doi:10.1128/AEM.00062-07

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

Werdell, P. J., and others. 2019. The plankton, aerosol, cloud, ocean ecosystem mission: Status, science, advances. Bull. Am. Meteorol. Soc. **100**: 1775–1794. doi:10.1175/BAMS-D-18-0056.1

Yeh, Y. C., J. McNichol, D. M. Needham, E. B. Fichot, L. Berdjeb, and J. A. Fuhrman. 2021. Comprehensive single-PCR 16S and 18S rRNA community analysis validated with mock communities, and estimation of sequencing bias against 18S. Environ. Microbiol. **23**: 3240–3250. doi:10.1111/1462-2920.15553

Yoon, G., I. Gaynanova, and C. L. Müller. 2019. Microbial networks in SPRING-semi-parametric rank-based correlation and partial correlation estimation for quantitative microbiome data. Front. Genet. **10**: 516. doi:10.3389/fgene.2019.00516

Zapata, M., S. Fraga, F. Rodríguez, and J. L. Garrido. 2012. Pigment-based chloroplast types in dinoflagellates. Mar. Ecol. Prog. Ser. **465**: 33–52. doi:10.3354/meps09879

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 Microbiol. Ecol. **52**: 79–92. doi:10.1016/j.femsec.2004.10.006

Acknowledgments

The authors thank the Plumes and Blooms team, especially Nathalie Guillocheau and Stuart Halewood, for their efforts collecting the data and samples analyzed here. The authors also thank Jeff Krause for sharing the DNA meta-barcoding samples analyzed here, and Tom Bell for sharing curated photosynthetically available radiation data. Plumes and Blooms is funded by the National Aeronautics and Space Administration (NASA; 80NSSC18K0735; 80NSSC21K1750), 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 D.C. 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). D.C. was also funded in part by a NASA Earth and Space Science Fellowship (Grant NNX16AO44HS02). Meta-barcoding samples were collected under the NASA Earth and Space Science Fellowship Program (Grant NNX12AO13H to E.K.W.) and National Science Foundation Award OCE-0850857 to C.A.C.

Submitted 14 October 2021
Revised 21 February 2022
Accepted 04 November 2022

Deputy editor: C. Elisa Schaum