



A global seasonal surface ocean climatology of phytoplankton types based on CHEMTAX analysis of HPLC pigments

Journal Article**Author(s):**

Swan, Chantal M.; Vogt, Meike; [Gruber, Nicolas](#) ; [Laufkötter, Charlotte](#) 

Publication date:

2016-03

Permanent link:

<https://doi.org/10.3929/ethz-b-000208709>

Rights / license:

[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International](#)

Originally published in:

Deep Sea Research Part I: Oceanographic Research Papers 109, <https://doi.org/10.1016/j.dsr.2015.12.002>

Funding acknowledgement:

153452 - SOGate: Phytoplankton ecosystem control of the Southern Ocean biogeochemical gate (SNF)

Author's Accepted Manuscript

A global seasonal surface ocean climatology of phytoplankton types based on CHEMTAX analysis of HPLC pigments

Chantal M. Swan, Meike Vogt, Nicolas Gruber, Charlotte Laufkoetter



PII: S0967-0637(15)30074-1
DOI: <http://dx.doi.org/10.1016/j.dsr.2015.12.002>
Reference: DSRI2558

To appear in: *Deep-Sea Research Part I*

Received date: 29 July 2015
Revised date: 1 December 2015
Accepted date: 2 December 2015

Cite this article as: Chantal M. Swan, Meike Vogt, Nicolas Gruber and Charlotte Laufkoetter, A global seasonal surface ocean climatology of phytoplankton types based on CHEMTAX analysis of HPLC pigments, *Deep-Sea Research Part I* <http://dx.doi.org/10.1016/j.dsr.2015.12.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A global seasonal surface ocean climatology of phytoplankton types based on CHEMTAX analysis of HPLC pigments

Chantal M. Swan^{a*}, Meike Vogt^a, Nicolas Gruber^a, Charlotte Laufkoetter^{a,b}

^a Institute for Biogeochemistry and Pollutant Dynamics, ETH Zürich, Switzerland

^b Present address: Princeton University, New Jersey, USA

*Corresponding author

Keywords: phytoplankton, global, climatology, HPLC, pigments, chemotaxonomy, CHEMTAX

ABSTRACT

Much advancement has been made in recent years in field data assimilation, remote sensing and ecosystem modeling, yet our global view of phytoplankton biogeography beyond chlorophyll biomass is still a cursory taxonomic picture with vast areas of the open ocean requiring field validations. High performance liquid chromatography (HPLC) pigment data combined with inverse methods offer an advantage over many other phytoplankton quantification measures by way of providing an immediate perspective of the whole phytoplankton community in a sample as a function of chlorophyll biomass. Historically, such chemotaxonomic analysis has been conducted mainly at local spatial and temporal scales in the ocean. Here, we apply a widely tested inverse approach, CHEMTAX, to a global climatology of pigment observations from HPLC. This study marks the first systematic and objective global application of CHEMTAX, yielding a seasonal climatology comprised of ~ 1500 $1^\circ \times 1^\circ$ global grid points of the major phytoplankton pigment types in the ocean characterizing cyanobacteria, haptophytes, chlorophytes, cryptophytes, dinoflagellates, and diatoms, with results validated against prior regional studies where possible. Key findings from this new global view of specific phytoplankton abundances from pigments are a) the large global proportion of marine haptophytes (comprising $32 \pm 5\%$ of total chlorophyll), whose biogeochemical functional roles are relatively unknown, and b) the contrasting spatial scales of complexity in global community structure that can be explained in part by regional oceanographic conditions. The results are publically accessible via

doi.pangaea.de/10.1594/PANGAEA.855412, and will guide future parameterizations of marine ecosystem models exploring the link between phytoplankton community structure and marine biogeochemical cycles.

1. Introduction

Marine phytoplankton form the base of the oceanic food web and account for roughly half of the global atmospheric carbon fixed annually (Field et al., 1998; Behrenfeld et al., 2006). The importance of quantifying the trophic activity of phytoplankton and its implications for carbon sequestration and global climate in past, present and future oceans continues to be a guiding tenet for numerous disciplines within marine and climate science (Falkowski, 2012). Moreover, the recent increase in the number of techniques for quantifying different types of phytoplankton using satellite observations (e.g., Alvain et al., 2008; Raitos et al., 2008; Bracher et al., 2009; Kostadinov et al., 2010; Brewin et al., 2010; Hirata et al., 2011; Torrecilla et al., 2011; Mustapha et al., 2014), as well as the categorization of phytoplankton by global ecosystem modelers into functional groups to differentiate their biogeochemical roles (Moore et al., 2002; Le Quéré et al., 2005), each highlight the value of a reliable means for synoptically capturing the composition of global algal communities. There are large spatial and temporal gaps in available field data with which to validate such models, which limits ability to characterize environmental drivers of phytoplankton biogeography and investigate how ecosystems will respond to future environmental change (Doney et al., 2009). These data gaps persist largely due to the discrepant scales of ocean basins versus typical ship-based ocean surveys, but also due to the often patchy as well as time-consuming nature of plankton identification research (Buitenhuis et al., 2013; de Vargas et al., 2015).

Among the methods to quantify global phytoplankton community composition from water samples, which include scanning electron microscopy, light microscopy, flow cytometry (Buitenhuis et al., 2012; 2013) and most recently genomics (de Vargas et al., 2015), the separation of phytoplankton pigments via High Performance Liquid Chromatography (HPLC) remains the fastest, most systematic, cost-effective, and quality-controlled method to date (van Heukelem and Hooker, 2011). In this study we apply chemotaxonomic analysis to a global dataset of climatological HPLC pigment concentrations (Peloquin et al., 2013) compiled as part of the Marine Ecosystem DATA (MAREDAT) initiative (Buitenhuis et al., 2013). Chemotaxonomy

concerns the translation of pigment concentrations into estimates of community composition as a function of the overall chlorophyll pool (Letelier et al., 1993; Mackey et al., 1996). In this way, HPLC data grant the opportunity for an immediate whole-community perspective in which to explore patterns of dominance among phytoplankton assemblages. This outlook is not afforded at the global scale by abundance data from microscopy, in part due to the selective nature of taxonomic information needs in past surveys, but also due to the limitation that picophytoplankton ($< 2 \mu\text{m}$ in size) evade detection by conventional microscopy (Sherrard et al., 2006). Analysis of HPLC pigments is thus of particularly great value over the vast tropical and subtropical ocean domains in which these smallest of autotrophs dominate (Chisholm et al., 1988; Campbell et al., 1994).

Coarse approximations such as diagnostic pigment analysis (Letelier et al., 1993; Claustre, 1994; Vidussi et al., 2001; Uitz et al., 2006), as well as more exhaustive inverse optimization schemes have been used to analyze HPLC pigment data (Mackey et al., 1996; Van der Meersch et al., 2008). In this study, we employ the most widely tested inverse approach, CHEMTAX (Mackey et al., 1996; Wright et al., 2009; 2010), which uses matrix factorization to optimize initial guesses of accessory pigment-to-chlorophyll a ratios of the phytoplankton classes assumed to be present in sample for a best fit to the bulk pigment measurements. Effective pigment analysis via CHEMTAX accordingly requires two *a priori* assumptions: 1) that the types of phytoplankton in a water sample and their accessory pigment-to-chlorophyll a ratios are known, and 2) that the accessory pigment-to-chlorophyll a ratios are constant across the data subset under consideration (i.e., that a homogeneous community of phytoplankton has been sampled). Pigment ratios vary widely in nature within a given phytoplankton class and even within given species due to differences in photoacclimation and nutrient availability, with the potential to confound CHEMTAX results (Latasa, 2007; Goericke and Montoya, 1998). With adequate sample numbers, this drawback can be minimized by organizing pigment data into subsets characterizing discrete light and nutrient regimes (e.g., by latitude or depth binning), or by cluster analysis to identify subsets with pigment ratio similarities representing discrete phytoplankton assemblages (e.g., Dandonneau and Niang, 2007; Torrecilla et al., 2011; Wolf et al., 2014).

To our knowledge, the present work signifies the first application of CHEMTAX to a global climatology of pigment data. As such, we were met with the

challenge of developing a systematic approach to data subdivision and pigment ratio matrix selection, both essential to obtaining valid estimates from CHEMTAX. We present here an objective approach to interpolation and clustering of the global climatological pigment data into subsets suitable for input to CHEMTAX, and subsequent initialization of the corresponding ratio matrices. The selection of pigment ratios for CHEMTAX has relied historically on choosing mean literature values from cultured or wild phytoplankton species (Wright and van den Enden, 2000; Gibb et al., 2001; Kozłowski et al., 2011; DiTullio et al., 2013). As a timely advantage, this study makes use of a recently published synthesis of open ocean pigment ratios for the major algal groups (Higgins et al., 2011), permitting our results to reflect updated phylogenetic classifications as well as the degree to which pigments are shared across taxa. The ratios correspond to taxonomic groupings operationally defined as “pigment types” (Wright et al., 2009; Higgins et al., 2011), a convention we maintain throughout the paper to emphasize the derivation of our result from HPLC data. An analogous comprehensive dataset on phytoplankton distributions from alternate field identification methods (e.g. microscopy) does not exist to quantitatively validate the CHEMTAX results on the global scale; therefore, where possible regional observations of phytoplankton assemblages from previously published studies are compared with the estimates herein.

Herein we present mixed layer seasonal estimates from CHEMTAX of 9 phytoplankton pigment types with distributions across the major ocean basins, and discuss the trophic and hydrographic associations with community composition. The $1^\circ \times 1^\circ$ degree seasonal climatological estimates from this study (see Appendices A.1 and A.2) have been made available as open-access data products in terms of relative or absolute fractions of total chlorophyll *a* for the pigment types at: doi.pangaea.de/10.1594/PANGAEA.855412. This comprehensive dataset of new taxonomic information derived from HPLC will ideally guide the parameterization and tuning of remote sensing and ecosystem models, opening a new pathway for investigating how phytoplankton biogeography both influences and responds to global to regional scale biogeochemistry.

2. Data and Methods

Among the first challenges in developing a consistent approach to analyzing the global pigment climatologies was maximizing the number of locations where phytoplankton community structure could be ascertained, as well as to establish an objective method for isolating homogeneous subsets of global pigment data for the application of CHEMTAX. In the following sections we describe the HPLC data availability, the training of a neural network to interpolate missing pigment concentrations, the use of cluster analysis to subdivide the global data set and last, initialization and evaluation of the CHEMTAX input.

2.1. HPLC pigment data

Global monthly climatological $1^\circ \times 1^\circ$ gridded HPLC pigment concentrations at all available depths from the MAREDAT archive were acquired from Pangaea: <http://doi.pangaea.de/10.1594/PANGAEA.793246> (Peloquin et al., 2013). Sampling locations in the data set span the Atlantic, Pacific and Southern oceans as well as the Mediterranean and Arabian seas; therefore, many of the major surface water masses and various trophic states of the ocean were sampled, including coastal and open ocean eutrophic, mesotrophic, and oligotrophic areas (see Peloquin et al., 2013, their Fig. 1). Concentrations for monovinyl chlorophyll *a* (chl_a), divinyl chlorophyll *a* (DVchl_a), and 7 frequently sampled accessory pigments: alloxanthin (allox), 19'-butanoyloxyfucoxanthin (but), chlorophyll *b* (chl_b), fucoxanthin (fuco), 19'-hexanoyloxyfucoxanthin (hex), peridinin (peri), zeaxanthin (zeax), were averaged over corresponding monthly long-term mean mixed layer depths from the National Oceanographic Data Center (www.esrl.noaa.gov/psd/data/gridded/data.nodc.woa94.html) yielding a matrix of 1,938 observations \times 9 pigments. Total chlorophyll in the gridded data set ranges 3 orders of magnitude from 0.01 mg/m^3 to 17.45 mg/m^3 . Concentrations for accessory pigments prasinoxanthin, violaxanthin and lutein, while available at some locations, were not used in the present analyses as they were significantly undersampled relative to the other pigments, thereby of limited use in the data interpolation.

2.2. HPLC data interpolation

To fully exploit the available HPLC information, missing accessory pigment concentrations were interpolated using pigment-to-chlorophyll relationships. Of the 1,938 MLD monthly mean samples, 442 observations had missing values for one or

more of the 7 accessory pigments. A neural network was constructed for interpolation of pigment data at these locations as it can operate on the nonlinear relationships between pigment ratios to reconstruct missing values, as accessory pigment to chlorophyll ratios covary within phytoplankton populations in the ocean (Dandonneau & Niang, 2007; Not et al., 2008). In order to express the total phytoplankton pool in terms of pigment concentration, total chlorophyll *a* concentration (Tchl_a) was computed as the sum of chl_a + DVchl_a, as the genus *Prochlorococcus* contains only the DVchl_a as its primary light harvesting pigment. Next, ratios of each accessory pigment to Tchl_a were computed to normalize pigment concentration to chlorophyll biomass. Of the 7 accessory pigments considered herein, allox and peri were on average an order of magnitude lower in their ratio to Tchl_a than the other pigments, as they are respectively native to only one phytoplankton taxon in the ocean (Table 1). Thus, to give equal weight to all pigments in training of the network for interpolation, pigment ratios were normalized through mapping from their original range to [-1 1] (Dandonneau and Niang, 2007).

	Allox	But	Chlb	Fuco	Hex	Peri	Zeax
Median	<0.01	0.07	0.06	0.10	0.24	0.02	0.13
Min.	0	0	0	0	0	0	0
Max.	0.43	0.71	1.35	2.35	1.74	0.51	2.00

Table 1. Median and ranges (prior to normalization) of accessory pigment-to-chlorophyll *a* ratios in the present dataset.

Matlab's Neural Network Toolbox® was used to train a self-organizing map (SOM) on the gridded monthly MLD mean pigment:Tchl_a ratios. The number of neurons contained within the SOM ($n = 38 \times 38$) was set approximately equal to the number of grid cells containing measured values for all 7 pigments. We evaluated several network sizes, but as the sole objective was to interpolate with information from topologically neighboring data, our approach was to identify all potential patterns of covariability among the pigment ratios (*sensu* Jouini et al., 2013). In the resultant SOM, the data were summarized by 80% of the neurons, and on average data values differed by less than 6% from the reference vector of their respective matching neuron, with no significant bias observed across any of the pigments. Grid cells containing 3 or fewer missing values for the 7 accessory pigments were presented to the network, and a missing value was replaced with its corresponding

neuron's reference value for the pigment ratio. The criterion to interpolate only grid cells with 3 or fewer missing values was set after constructing four discrete 50 sample x 7 accessory pigment synthetic data subsets following Latasa (2007), applying a cluster analysis to the normalized pigment:Tchl_a ratios of the synthetic data, and assessing the effect of missing data on the clustering. In brief, a missing value for a randomly chosen pigment was inserted stepwise into each sample's pigment suite in the synthetic dataset until the additive effect of the missing data no longer resulted in the synthetic data clustering into its four original subsets. This threshold condition was reached when the number of missing pigment values in a synthetic data sample exceeded 3. In other words, samples containing measured values for at least 4 of the 7 accessory pigments were considered to be suitably informative for pattern recognition, thus eligible for interpolation. All but 15 samples met this criterion, thus overall the SOM-based interpolation increased the number of monthly MLD mean grid cells with a complete pigment suite from 1,496 to 1,923 while conserving the distribution functions of pigment concentration (not shown).

2.3. Clustering of HPLC data

Seasonal means of the interpolated monthly MLD pigment ratios were computed for the northern hemisphere (where "winter" = Jan - Mar., "spring" = Apr.-Jun., "summer" = Jul. - Sept., "autumn" = Oct. - Dec.) and southern hemisphere (where "winter" = Jul. - Sept., "spring" = Oct. - Dec., "summer" = Jan. - Mar., "autumn" = Apr. - Jun.) reducing the data matrix to 1,667 observations. A single hierarchical cluster analysis was performed on the global dataset of seasonal MLD mean accessory pigment ratios using the Euclidean distance metric and Ward's linkage criterion (Figs. 1 and 2; Peng et al., 2011). In contrast to the SOM approach used for interpolation (section 2.2), pigment ratios were not normalized for the cluster analysis so as not to give undue weight to minor accessory pigments (e.g., allox, peri and chl_b; Table 1); therefore, the clustering was primarily forced by patterns among the major accessory pigments (fuco, hex and zeax; Table 1). Minimization of the Davies-Bouldin index (Davies and Bouldin, 1979), based on the ratio of within-cluster to between-cluster distances and evaluated for 2 through 50 clusters, indicated an optimal separation of the data into 6 clusters (Fig. 2, inset).

The clusters demonstrate coherence with the geographic distribution of the data, as seen on a global map (Fig.1) and by their relative proximities on a

dendrogram (Fig. 2), in which data approximately organize into clusters of tropical and subtropical locations (clusters 3 and 4), polar locations (clusters 5), and the eastern section of the Antarctic coast (clusters 1 and 2). Only one of the clusters (cluster 6) has a relatively undefined regionality (Fig. 1). The clusters were not confined to any particularly season (with the exception of the bulk of cluster 5 data occurring in summer due to a temporal bias in sampling conducted in the Southern Ocean; Fig. 1). A non-metric multidimensional scaling analysis (Borg and Groenen, 2005) applied to the Euclidean distance matrix from the clustering illustrated a clear and tight separation of the data clusters with virtually no overlap between clusters in 7-dimensional pigment ratio space (Supplementary Fig. S1). The 6 clusters were therefore taken to represent data subsets with sufficiently distinct pigment ratio profiles for analysis by CHEMTAX.

FIGURE 1

FIGURE 2

2.4. Chemotaxonomic analysis of HPLC data

2.4.1. Initialization of CHEMTAX

To complete the objective of isolating pigment data subsets comprising a discrete phytoplankton community, each data cluster was assumed to be qualitatively comprised of the same community of pigment types; however, each season of each cluster was run separately in CHEMTAX to allow for potential temporal community shifts in relative phytoplankton abundance.

Solving for estimated taxonomic abundance via CHEMTAX requires that the number of taxonomic groups not exceed the number of pigments (including Tchla) considered in the calculation, so as to prevent linear dependency of the initial pigment ratio matrix (Mackey et al., 1996). This implied a limit of 7 pigment types that could be included in the initial ratio matrices as 7 accessory pigments and the pigment chlorophyll were considered.

Work in recent years has uncovered that most all pigments, with few exceptions, are shared across multiple phytoplankton taxa (Zapata et al., 2004; see Jeffrey et al. 2011 for a complete summary). A select few of the 7 accessory pigments

in the data set are unambiguously diagnostic of the presence of certain phytoplankton groups: allox (cryptophytes), peri (*Amphidinium* sp. dinoflagellates) and DVchla (*Prochlorococcus* sp.). *Prochlorococcus* abundance can be directly determined from DVchla concentration, so a pigment type representing this genus was not explicitly included in the initial ratio matrix. Based on the available HPLC data, the pigment types chosen for evaluation by CHEMTAX herein (and the representative genus of each pigment type suggested by Higgins et al., 2011) were: CYANO-1 (cyanobacteria, e.g. *Trichodesmium* sp.), CYANO-2 (cyanobacteria, e.g. *Synechococcus* sp), which initially includes the implicit contribution by *Prochlorococcus* sp. (“CYANO-4”) that may be corrected for *a posteriori* by subtracting the DVchla concentration; DIATOM-1 (diatoms, e.g. *Chaetoceros* sp.), HAPTO-6 (haptophytes, e.g. *Emiliania huxleyi*), HAPTO-8 (haptophytes, e.g. *Phaeocystis* sp.), CRYPTO-1 (cryptophytes, identifiable by presence of allox), DINO-1 (dinoflagellates, e.g. *Amphidinium* sp., identifiable by presence of peri), and CHLORO-1 (chlorophytes, otherwise known as green algae, e.g. *Dunaliella* sp.). The ubiquity of chl_b in the global dataset signals the potential presence of chl_b-containing prasinophytes and/or euglenophytes in addition to green algae. However, as green algae are more abundant in the open ocean than the other chl_b-containing taxa, we opted to include CHLORO-1 in the pigment ratio matrix in favor of pigment types characterizing open ocean prasinophytes or the less-frequently occurring euglenophytes (Andersen et al., 1996; Not et al., 2008). Resulting CHLORO-1 fractions thus contain implicit pigment contributions by other chl_b-containing taxa. Chl_b is known to co-occur with zeax in varying proportions in most all known chl_b-containing algae (Schlüter and Mohlenberg, 2003), so the consistent inclusion of CHLORO-1 in the ratio matrix also ensured that zeax was not overallocated to cyanobacteria globally, and could be allocated appropriately to green algae in the high latitudes where, as mentioned, cyanobacteria do not constitute a major group (Wright et al., 2000; Zwirgmaier et al., 2008)

Selection of appropriate pigment ratios to characterize the above pigment types was necessary as CHEMTAX is sensitive to these initial guesses (Latasa, 2007). Published summary ratio tables for each pigment type, which include ratios for ‘low-‘ (LL), ‘medium-‘ (ML) and ‘high-light’ (HL) environments (Higgins et al., 2011), were utilized for this purpose. As light availability is a fundamental determinant of intracellular pigment variability, we made an assessment of the median mixed layer

light level exemplifying each season of each data cluster using the global MLD (section 2.1), as well as surface photosynthetically available radiation (PAR, Ein m⁻² d⁻¹) and the diffuse attenuation coefficient for downwelling irradiance at 490 nm (k_d490 , m⁻¹) from SeaWiFS following Behrenfeld et al. (2005), with threshold PAR ranges for low, medium and high light environments defined as <15 Ein m⁻² d⁻¹, 15 – 35 Ein m⁻² d⁻¹, and >35 Ein m⁻² d⁻¹, respectively. The field ratio at the light level best characterizing the majority (at least 65%) of data locations per cluster per season was selected for input.

Confidence in the results derived from CHEMTAX, thus confidence in the choice of pigment types to include in the initial ratio matrix, is obtained when both the residual pigment content from the fits is low across the samples, and when the individual pigment ratios converge to within known ranges. In this manner, CHEMTAX is suited to hypothesis testing by modifying the initial phytoplankton groups evaluated (Wright et al., 2009; 2010). As a starting point, and based on the presence of their diagnostic pigments in the dataset, HAPTO-6 (hex, fuco), HAPTO-8 (hex, fuco), CHLORO-1 (chl_b, zeax), CRYPTO-1 (allox), DINO-1 (peri), and DIATOM-1 (fuco) were all included in the initial ratio matrix for each cluster. Building on this selection, CYANO-1 and CYANO-2 were initially included in clusters in which 65% or more of the data were located between 40°S and 40°N based on the typically low cyanobacterial abundance observed poleward of these latitudes (Zwirgmaier et al., 2008). Several replacements of pigment types from the ratio matrices for the clusters had to be made during initial trial runs in CHEMTAX until the combination of pigment types resulted in low residual values and final pigment ratios within natural ranges for each cluster (Higgins et al., 2011). We describe these replacements in detail next in section 2.4.2.

2.4.2. Evaluation of CHEMTAX

Original CHEMTAX source code for Matlab (Mackey et al., 1996) was obtained from S. Wright (Australian Antarctic Division, Tasmania). Initial ratio matrices were iteratively modified in a minimization scheme to determine a best fit to the pigment data, with output ratios used to allocate the portion of chlorophyll *a* attributed to each pigment type. As CHEMTAX attempts to solve an overdetermined system of equations, we adapted a randomized trial strategy for the input ratio matrices to avoid local minimum solution space and thus divergence of the ratios

during the optimization procedure. To simulate deviations from the pigment ratio matrix due to the natural variations in phytoplankton pigment content, each non-zero value in the initial ratio matrix was randomly varied up to $\pm 35\%$ and the minimization was repeated 60 times following Wright et al. (2009). In recent reports on studies employing CHEMTAX, either the solution with the lowest residual root mean square, or the mean of the 10% of results ($n = 6$) with the lowest residuals, is used to estimate the chlorophyll concentration of each phytoplankton group (e.g., Wright et al., 2009; 2010; Schlüter et al. 2011; Kozłowski et al., 2011). Using the synthetic data discussed in section 2.2, a quality check on the iterative method just described returned final chlorophyll values (determined as the mean of the top 10% of results) that on average fell within $\pm 6\%$ of a sample's "true" chlorophyll value for each pigment type, leading us to choose this approach to produce the final taxonomic estimates.

There is no established measure of uncertainty for CHEMTAX applied at the global scale. Wright and Jeffrey (2006) indicate that taxonomic groupings estimated from pigment concentrations can be expected to have greater uncertainty than that of pigment data themselves. Considering our synthetic data test using CHEMTAX, and given the average inherent analytical uncertainty in the HPLC dataset of at least 7.4% (Van Heukelem & Hooker, 2011; Kozłowski et al., 2011), we suggest a conservative bound for uncertainty of $\pm 15\%$ of the absolute and relative Tchl_a fractions from CHEMTAX reported herein.

As mentioned, a few adjustments of pigment types to some of the initial ratio matrices were made until best fits to the data were achieved. Clusters 3, 4 and 6 initially included CYANO-1, but with the exception of cluster 3 and cluster 4 in winter, inclusion of CYANO-1 yielded abnormally high final zeax ratios for either CHLORO-1 or CYANO-2, and was thus removed from the matrix in such cases. This was likely due to CYANO-1 (e.g., *Trichodesmium* sp.) representing a minor group (i.e., less than 5% percent chlorophyll) at most locations (Capone et al., 1997; Subramaniam et al., 2002; Carpenter et al., 2004), as CHEMTAX is not optimized to correctly resolve minor groups that share pigments with major groups (Mackey et al., 1996; Latasa, 2007). CRYPTO-1 was ultimately removed from the ratio matrices for clusters 3 and 4 as allox concentrations were observed to be consistently low (<0.001 mg/m³ on average). Finally, we experimented with elimination of HAPTO-6 versus HAPTO-8 from the initial ratio matrices of cluster 3 and cluster 4 in winter, as hex

ratios often diverged to values above maximum field estimates when both types were simultaneously included. Ultimately for these cases, hex ratios converged to known ranges when HAPTO-6 was included in favor of HAPTO-8.

The final pigment type assemblages that demonstrated best fits for each cluster are shown in Table 2 and were used to produce the final ratio matrices, which represent the mean of ratios from the top 10% results with lowest residuals. In the top 10% of results, consistently <1% of the chlorophyll content for each cluster went unexplained by CHEMTAX. Both initial and final pigment ratios for each season of each cluster are provided in Appendices A.3 and A.4, respectively.

Cluster #	CYANO-1	CYANO-2	DIATOM-1	HAPTO-6	HAPTO-8	DINO-1	CRYPTO-1	CHLORO-1
1			❖	❖	❖	❖	❖	❖
2			❖	❖	❖	❖	❖	❖
3	❖	❖	❖	❖		❖		❖
4 Winter	❖	❖	❖	❖		❖		❖
4		❖	❖	❖	❖	❖		❖
5			❖	❖	❖	❖	❖	❖
6		❖	❖	❖	❖	❖	❖	❖

Table 2. Pigment types in each cluster for which final seasonal abundance estimates from CHEMTAX are reported herein.

Values of zero for relative and absolute chlorophyll were assigned to any of the above pigment types that were deliberately excluded from the CHEMTAX calculation as their omission from the ratio matrix effectively implied their absence. The final CHEMTAX estimates of CYANO-2 were corrected for the contribution of *Prochlorococcus* by subtracting the DVchl_a concentration from the CYANO-2 Tchla concentration, while the DVchl_a concentration subsequently supplied the estimate for CYANO-4 (*Prochlorococcus*). For a total of 146 grid cells in the seasonal climatologies we observed negative relative Tchla values for CYANO-2 after this correction, averaging -8% and ranging up to -92%. This suggested a source of error potentially related to the CHEMTAX determination, raw DVchl_a measurement, and/or the geographic averaging of the raw DVchl_a measurement. Extreme negative values (defined as < -10% Tchla) were observed in only 56 of the grid cells. These cases were considered outliers, and values for all pigment types were replaced with missing values. Where the deviations were small (ranging from near zero to -10%) it was assumed that the CYANO-2 CHEMTAX determination was entirely attributed to *Prochlorococcus*, and the CYANO-4 value estimated from DVchl_a was replaced by

the CHEMTAX estimate for CYANO-2 (while CYANO-2 was replaced with null value) in order to conserve total relative chlorophyll at 100%. Final CYANO-2 estimates thus reflect contribution by *Synechococcus*-like cyanobacteria, while CYANO-4 estimates reflect that of *Prochlorococcus*.

2.4. Environmental data

We evaluated the relative abundance of each pigment type against several properties of or influencing phytoplankton growth in the mixed layer. These properties include mixed layer mean temperature ($^{\circ}\text{C}$) and nitrate concentration (μM) from the World Ocean Atlas (<https://www.nodc.noaa.gov/OC5/woa13/>), as well as the median mixed layer light level ($\text{Ein m}^{-2} \text{d}^{-1}$) computed using MLD climatologies and SeaWiFS monthly climatologies for PAR and k_490 following Behrenfeld et al. (2005) as in section 2.4.1. Also included is the surface wind stress magnitude (N m^{-2}) and Ekman upwelling velocity (m d^{-1}) calculated using monthly climatologies for wind stress curl and zonal wind stress from the Scatterometer Climatology of Ocean Winds (<http://www.locean-ipsl.upmc.fr/~cdblod/mld.html>) following Risien and Chelton (2008). Products were gridded to $1^{\circ} \times 1^{\circ}$ and seasonally averaged to match the scales of the pigment data.

3. Results and Discussion

A number of studies have already employed CHEMTAX or diagnostic pigment equations (e.g., Letelier et al. 1993; Uitz et al. 2006) to analyze several regional subsets of the raw pigment dataset from which the gridded pigment climatologies herein derive (e.g. Aiken et al. 2009; Goericke, 2002; Wright and van den Enden, 2000; Wright et al., 2009; 2010; Schlüter et al., 2011; Kozłowski et al. 2011), wherein HPLC-derived taxonomy was validated using associated microscopic data. A co-located dataset of, for example, microscopy observations to validate the CHEMTAX results on the global scale does not exist at present, and measurements from light microscopy alone do not permit comparisons for areas such as the oligotrophic subtropics where picoplankton dominate 80% or more of the chlorophyll pool. Therefore, where possible we compare the resultant distributions of the pigment types evaluated herein with literature reports in which higher resolution regional assessments have been made using HPLC, microscopy and/or flow cytometry data.

Below we first describe the characteristics of the data clusters, the construct within which CHEMTAX was applied. We then provide a descriptive account of global community structure highlighting distinctions as well as commonalities among a number of well-sampled Longhurst provinces within the dataset (Longhurst, 2007). Community structure is thereafter regressed against environmental conditions, and finally, phytoplankton specific biomass estimates are given for the global ocean.

3.1. Distribution and character of the data clusters

3.1.1. Geographic characteristics of the clusters

The majority (nearly 93%) of the data are contained within clusters 4, 5 and 6 (Fig. 2; Table 3). Less than 3% of the dataset comprises cluster 1 (“E. Antarctic Shelf high relative fuco” cluster, n = 16) and cluster 2 (“E. Antarctic Shelf” cluster, n = 27), each confined to polar shelf waters within the Indian sector of the Antarctic. Cluster 3 (“Atlantic Sub/tropical”, n = 77) also contains relatively few data (< 5% of total), occurring primarily in the South Atlantic subtropical gyre and equatorial region, with a few points in the tropical Pacific and Arabian Sea. Cluster 4 (“Global Sub/tropical” cluster, n = 574) encompasses the majority (75%) of remaining open ocean data within the subtropical gyres and equatorial regions. Cluster 5 (“Polar” cluster, n = 508) is characterized primarily by data in the polar and subpolar regions (with a sampling bias toward the Southern Ocean), while containing a fraction of coastal data from all seasons (n = 89), as well as a few data from the eastern North Atlantic in spring (n = 19). Finally, Cluster 6 (“Mesotrophic” cluster, n = 465), which is the least constrained geographically, has a global distribution in relatively nutrient-replete waters near the continents, including: waters within the Denmark Strait, waters in and neighboring the Brazil-Malvinas confluence zone as well as the transition zone between the Chilean upwelling and the South Pacific gyre, several locations in the Benguela upwelling system, Pacific equatorial upwelling, and locations in the eastern North Atlantic, Mediterranean and Arabian Seas (refer to Fig. 1).

Cluster #	Description	Winter	Spring	Summer	Autumn
1	E. Antarctic Shelf high relative fuco	n=0	n=5	n=11	n=0
2	E. Antarctic Shelf	n=4	n=7	n=16	n=0

3	Atlantic Sub/tropical	n=3	n=23	n=8	n=43
4	Global Sub/tropical	n=67	n=185	n=163	n=159
5	Polar	n=68	n=133	n=252	n=55
6	Mesotrophic	n=127	n=110	n=112	n=116

Table 3. Number of seasonal data (mixed layer mean pigment suites) contained per cluster.

3.1.2. Pigment characteristics of the clusters

Based on the bulk pigment ratio medians characterizing each cluster, zeax forces the primary division of the data, in which clusters 3 and 4 oppose clusters 1, 2, 5, and 6 (Figs. 2 and 3). Clusters 1, 2 and 3, containing less than 8% of the data, exhibit the highest median values for fuco (1.70), hex (1.06) and zeax (1.10) ratios, respectively, relative to other pigments of the cluster, as well as relative to all clusters. Clusters 4, 5 and 6, containing the remaining data, are distinguished, respectively, by values of zeax (0.50), fuco (0.41) and hex (0.37) that are high within the cluster relative to other pigments, although lower overall than the zeax, fuco and hex values distinguishing clusters 1, 2 and 3.

FIGURE 3

Clusters 1 and 2 inhabit a limited region of East Antarctic coastline and exhibit atypically high fuco (up to 2.4) and hex (up to 1.7) ratios with respect to the global distribution. Clusters 1 and 2 are distinct from each other in that the fuco ratio in cluster 1 exceeds that of hex by four-fold, whereas in cluster 2 this trend is reversed albeit to a lesser extent (hex \approx 1.5x fuco). The high relative fuco ratio of cluster 1 is explained by the low chlorophyll content relative to other clusters (discussed next in section 3.2). Similarly, the high relative fuco and hex ratios of cluster 2, and the high relative zeax ratio distinguishing cluster 3 are also consequences of low chlorophyll content in relation to clusters 4, 5 and 6 (also discussed in 3.2). Next, we discuss these defining pigment characteristics in light of the CHEMTAX results for the clusters.

3.1.2. Community structure of the clusters

The annual median community structure (Fig. 4) illustrating the most abundant pigment types in each cluster calls to mind the pigment ratio profiles that distinguished the clusters (Fig. 3). (A pigment type is referred to herein as ‘most abundant’ if comprising at least 30% Tchl_a. In cases where none or more than one pigment type met this definition, or where contributions in excess of 30% Tchl_a were equal among two or more pigment types, we refer to a ‘mixed assemblage’.) DIATOM-1 is the most abundant pigment type in clusters 1 and 5, reflective of the high relative fuco in these clusters. Similarly, the relative high abundances of DIATOM-1 and HAPTO-6 in cluster 2 reflect its high relative fuco and hex values, while the most abundant pigment type in cluster 6, HAPTO-6, is consistent with its high hex ratio. The elevated zeax ratio of clusters 3 and 4 is manifest in CYANO-4 being most abundant in these clusters (Fig. 4).

FIGURE 4

Averaging across clusters, DINO-1 and CRYPTO-1 do not contribute more than 5% to the total chlorophyll pool. The low relative abundance of the single pigment markers peri and allox (Fig. 3), respectively, suggest that DINO-1 and CRYPTO-1 represent minor groups in the ocean. DINO-1 in particular represents only the relative small marine fraction of dinoflagellate genera containing peridinin. Particularly within some areas of the coastal biome, blooms of various non-peridinin-containing dinoflagellates may constitute up to 100% of the autotrophic biomass (Smayda, 2002); however, such taxa could not be explicitly resolved using the present dataset. Rather, due to the pigments of their endosymbionts, contributions by a variety of dinoflagellate taxa are inherent in the DIATOM-1, HAPTO and CHLORO-1 fractions determined herein (Jeffrey et al., 2011; Higgins et al., 2011).

The distinction between communities of clusters 1 and 2 is mainly described by the relative high abundance of diatoms in cluster 1 versus the high abundance of both haptophytes and diatoms in cluster 2 (Fig. 4), which is supported by a detailed chemotaxonomic evaluation by Wright and van den Enden (2000) of the East Antarctic shelf between 80°E and 150°E, over which clusters 1 and 2 are distributed. In that study, sites with stronger thermal stratification along this stretch of Antarctic coastline exhibited high concentrations of diatoms, while locations with well-mixed layers were linked with relative high haptophyte abundance. Adding to this,

microscopic analyses suggested that associated zonal variation in heterotrophic grazer composition also contributed to the regional differences in relative algal abundance that we observe between clusters 1 and 2 (Wright and van den Enden, 2000).

The average community of cluster 1 looks similar to that of cluster 5 when viewed as a function of relative chlorophyll (Fig. 4, upper panel), while inspection of the absolute chlorophyll concentration reveals that cluster 1 is characterized by an order of magnitude lower DIATOM-1 concentration than cluster 5 (Fig. 4, lower panel). Although data within these clusters occur within the same region, and the CHEMTAX analysis reveals roughly the same relative community structure in each, the cluster analysis was influenced by the differential effect of chlorophyll concentration on the pigment ratios, thus permitting separate data treatments (CHEMTAX optimizations). Data in clusters 1 and 2 originate from Antarctic surveys conducted in austral summer of 1996 (Wright and van den Enden, 2000), when maximum and integrated chlorophyll stocks were roughly one-third and one-half the levels, respectively, of those measured in the contiguous marginal ice zone between 30°E and 80°E (i.e., measurements contained in cluster 5) one decade later (Wright et al., 2010), which explains our observed differences in absolute chlorophyll between clusters 1 and 5. The difference in chlorophyll level was partially attributed to interannual differences in sea-ice retreat as it influenced bloom timing (Wright et al., 2010). This highlights an example where sampling biases inherent in the global pigment dataset influence its climatological interpretation.

Clusters 3 and 4 also exhibit geographic overlap. The difference in relative zeax content between clusters 3 and 4 can be attributed to interannual differences in zeax and/or chlorophyll reported along Atlantic Meridional Transect (AMT) cruises between 1995 and 2004, which are the source of the South Atlantic observations in clusters 3 and 4 (Aiken et al., 2009; Peloquin et al., 2013). The primary feature distinguishing the annual median community of clusters 3 and 4 is the inclusion of the CYANO-1 fraction in cluster 3. This pigment type is purported to represent the nitrogen-fixing cyanobacterium *Trichodesmium*, abundances of which were microscopically confirmed during the AMT campaign, albeit with highest abundances recorded between 5°N – 15°N (Tyrrell et al., 2003; Luo et al., 2012), which we do not observe in the current data set. In general, this genus dominates in the North Atlantic and its observation can be sporadic and limited to when it forms colonial mats during strong blooms (Carpenter et al., 2004). CYANO-1 is in appreciable concentration in

cluster 3, ranging up to 53% of Tchla, yet is not found to be a significant regional or global contributor to chlorophyll biomass in the present climatology overall. Since the median annual community for cluster 3 was computed from 66 observations as compared to the 574 observations used to compute that for cluster 4, results from cluster 4 are more likely to be representative of the average surface community structure in the subtropical South Atlantic.

With regard to the variability about the cluster median communities seen in figure 4, the widest range in relative Tchla for a given pigment type occurs in the cluster in which it is most abundant (Table 4). For example, in clusters 1, 3 and 5 HAPTO-6 ranges to approximately one-third of total chlorophyll, whereas in clusters 2 and 6 its range is two-thirds or more. CYANO-2 and CHLORO-1 Tchla contributions are constrained to within 23% in cluster 3, significantly lower than their ranges in the other clusters, and a large discrepancy in the range of CRYPTO-1 (11% versus 67% Tchla) is observed between clusters 1 and 2 despite data from these clusters occupying the same Antarctic coastal region.

	Cluster 1 (n=16)	Cluster 2 (n=27)	Cluster 3 (n=77)	Cluster 4 (n=574)	Cluster 5 (n=508)	Cluster 6 (n=465)
CYANO-4	-	-	18 – 58%	0 – 81%	-	0 – 39%
CYANO-2	-	-	0 – 23%	0 – 82%	-	0 – 48%
HAPTO-6	0 – 31%	0 – 69%	15 – 37%	0 – 42%	0 – 38%	0 – 97%
HAPTO-8	3 – 39%	0 – 71%	-	0 – 38%	0 – 69%	0 – 54%
CHLORO-1	1 – 47%	0 – 67%	0 – 23%	0 – 50%	0 – 77%	0 – 70%
CRYPTO-1	0 – 11%	0 – 67%	-	-	0 – 98%	0 – 47%
DINO-1	0 – 5%	0 – 9%	0 – 12%	0 – 20%	0 – 50%	0 – 30%
DIATOM-1	15 – 80%	0 – 61%	0 – 11%	0 – 45%	0 – 96%	0 – 26%
CYANO-1	-	-	0 – 42%	0 – 53%	-	-

Table 4. Range in % Tchla for each pigment type per cluster

While overlap in the ranges of variability among individual pigment types both within and between clusters is observed, difference in ranges of each pigment type across clusters (Table 4) suggest underlying community structure patterns distinct to each cluster that extend beyond the predominant pigment type. To test this hypothesis statistically, as well as assess how scales of between-cluster variability compare to within-cluster variability, a non-metric multidimensional scaling analysis

(NDMS) was applied to the relative fractions of pigment types in all data points (Fig. 5, in which colored markers correspond to the 6 clusters). Although some overlap is observed most noticeably between clusters 5 and 6, the major clusters 4, 5 and 6 are statistically distinct from one another in terms of community structure. The NMDS ordination also reveals that the distances between points within a cluster span the same distances as those between clusters, although diverging in different directions. This suggests that within-cluster dissimilarities in community structure is on the same order of magnitude as that between clusters, even so the communities of each cluster are statistically different from one another.

FIGURE 5

To further evaluate which pigment types drive the distinctions in community structure observed between the major clusters, non-parametric ANOVAs by ranks (Kruskal-Wallis test) and *post hoc* pair-wise multiple comparisons (Dunn's tests) were performed for each pigment type in clusters 4, 5 and 6. Not surprisingly, significant differences were observed between each pair of clusters for the predominant pigment types DIATOM-1, HAPTO-6, CYANO-2 and CYANO-4 ($p < 0.05$). No distinctions between clusters could be made on the basis of DINO-1 ($p = 0.38 - 1.00$). However, HAPTO-8 fractions were significantly different between clusters 4, 5 and 6 ($p < 0.05$). Further, CHLORO-1, CRYPTO-1 and CYANO-1 fractions were significantly different between clusters 4 and 5 ($p < 0.05$) and clusters 4 and 6 ($p < 0.05$), respectively, but not between clusters 5 and 6 ($p = 0.97$ for CHLORO-1, $p = 0.08$ for CRYPTO-1, $p = 1.00$ for CYANO-1). This result supports our hypothesis that individual contributions of HAPTO-8, CHLORO-1, CRYPTO-1, and CYANO-1 (in addition to DIATOM-1, HAPTO-6, CYANO-2 and CYANO-4) each play a significant role in determining the relative community structure of clusters 4, 5 and 6 in which most of the global data are contained.

3.2 Community structure in the sampled Longhurst provinces

In order to further assess the spatial and temporal variability in the phytoplankton assemblage about the median community of each cluster (Fig. 4) in relation to regional oceanography, we subdivided the major data clusters 4, 5 and 6 according to Longhurst province definitions (Longhurst, 2007), whereby a minimum

of 10 samples was required to compute the median community structure for the province (Fig. 6). We did not apply the same treatment to clusters 1, 2, or 3 as with the exception of a few samples, data in these clusters were situated, respectively, within a single province; i.e., the Antarctic Polar province (APLR) for clusters 1 & 2, and the South Atlantic Gyre province (SATL) for cluster 3, hence the cluster median community (Fig. 5) represents the province level median community for these clusters. (Hypotheses for why clusters 1, 2 and 3 differ from the major clusters 4, 5 and 6 were given in section 3.1.2.)

FIGURE 6

It is evident from the distribution of clusters that data from more than one cluster are on occasion situated in the same region during the same season (refer to Fig. 1). Accordingly, a striking observation in community structure at the province level (Fig. 6) is the occurrence of the Northwest Arabian Upwelling (ARAB), eastern North Atlantic subtropical gyre (NASE), North Atlantic Drift (NADR), Antarctic (ANTA) and Austral Polar (APLR) provinces in multiple clusters during the same season, indicating several distinct communities existing within a season in one province. (This also indicates that the geographic overlap of clusters in some regions does not necessarily translate to instances of overlap in community structure of clusters as seen from the NMDS analysis in section 3.1.2.) Data in the ARAB province is contained in clusters 5 and 6 in winter, in clusters 4, 5 and 6 in summer, and clusters 4 and 6 in autumn. Data in the NASE and NADR provinces occur in clusters 4, 5 and 6 in spring, while the ANTA and APLR provinces are sampled in clusters 5 and 6 in summer (Fig. 6). In the following subsections we relate observed diversity in the phytoplankton community to the regional oceanographic characteristics of the aforementioned well-sampled provinces. We then further contrast this with basin scale patterns in community structure across provinces within oligotrophic regimes.

3.2.1 Antarctic Provinces

The two separate communities determined for the ANTA and APLR provinces in summer (seen in clusters 5 and 6; Fig. 6) can be explained by zonal biogeography in the circumpolar Southern Ocean. The samples from ANTA and APLR in cluster 6

in summer are located primarily in the Ross Sea and dominated by haptophytes, whereas the majority of remaining data in the ANTA and APLR provinces is contained within the diatom-dominated cluster 5 extending from the Antarctic Peninsula and eastward.

The abundance of diatoms in summer in ANTA and APLR provinces extending eastward from the Antarctic Peninsula is well documented (Wright and van den Enden, 2000; Wright et al., 2010; Kozłowski et al., 2011), and equally supported by the present analysis. In fact, the CHEMTAX optimization returned final pigment ratios for diatoms, dinoflagellates, cryptophytes, chlorophytes and haptophytes in the East Antarctic in remarkable agreement (average difference of $\pm 22\%$ of final ratios) with those reported by Wright and van den Enden (2000) and Wright et al. (2010) for this region (see Appendix A.4). This is noteworthy when considering that updated pigment ratio information as well as coarser data resolution is used herein compared to prior CHEMTAX evaluations in this region. In fact, even interannual variability in dominance of diatoms versus cryptophytes along the Antarctic Peninsula, described previously by Mendes et al. (2013), is captured by the present CHEMTAX analysis. This is illustrated on the seasonal composite map depicting the most abundant pigment type in each sample (Fig. 7).

Overall, the climatological seasonal variation along the Antarctic Peninsula (Figs. 6 and 7), wherein abundances of DIATOM-1 from winter through summer (APLR; cluster 5) are succeeded by a mixed community of DIATOM-1, CHLORO-1, HAPTO-8 and HAPTO-6 in autumn (APLR; cluster 6) are highly consistent with succession patterns observed in the multi-year time series at Palmer Station (Smith et al., 1998), with the CHLORO-1 fraction of all Antarctic samples herein largely attributable to chl b -containing prasinophytes (Kozłowski et al., 2011; Wright et al., 2010).

FIGURE 7

In contrast to the diatom dominance observed throughout much of the surveyed Southern Ocean (Wright et al., 2010; Kozłowski et al., 2011), the Ross Sea sector of the ANTA and APLR provinces is considered exceptional with respect to Southern Ocean waters in that it is especially dominated by the haptophyte species *Phaeocystis antarctica* over spring and summer (Arrigo et al., 1998). As there are few

documented occurrences of coccolithophore abundance south of the Antarctic polar front (Gravalosa et al., 2008; Wolf et al., 2014), it is likely that the HAPTO-6 abundances we observe in the Ross Sea are tracking iron-limited variants of *P. antarctica*, which exhibit highly variable pigment content in response to light and iron stress (Smith et al., 2010; Wright et al., 2010), with colonial strains in the Ross Sea displaying significantly higher hex ratios than other global *Phaeocystis* species (van Leeuwe and Stefels, 2007; DiTullio et al., 2013). These ratios are in fact closer to the initial ranges of HAPTO-6 hex ratios herein (see Appendix A.3), which explains why data from the Ross Sea were grouped into cluster 6, as it is characterized by high relative hex (Fig. 3). By contrast, HAPTO-8 estimates (typically representing *Phaeocystis*) were on average close to zero in this region. The Ross Sea thus represents an area of uncertainty in our estimates, where regional specificity in algal physiology eluded correct identification via CHEMTAX. At the same time, the CHEMTAX result, combined with prior knowledge on the low likelihood of coccolithophores in the region, is useful in confirming this known iron-limited *P. antarctica* regime. Less than 2% of the HPLC data is located in the Ross Sea, thus the above considerations do not affect our overall interpretations of global haptophyte abundance. However, future assessments utilizing results from the present study will be potentially more accurate if the HAPTO-6 estimates in the Ross Sea are considered to represent the local *P. antarctica* community (and not coccolithophores as HAPTO-6 may represent).

3.2.2. Arabian Sea Province

The ARAB province is well sampled in the present dataset, as it is an interest region of high biological diversity owing to frequent mesoscale features in the general circulation (e.g., Garrison et al., 1998; Latasa and Bidigare, 1998; Roy et al., 2015). The pigment climatologies for the Arabian Sea herein derive from 1995 JGOFS process cruises in this region (Peloquin et al., 2013 and references therein). The seasonal composites of dominance (Fig. 7) also illustrate the proximity of samples with different dominant pigment types within the Arabian Sea in a given season.

The Arabian Sea is a seemingly confined area to host several phytoplankton assemblages within a given season as implied by the multiple communities determined by CHEMTAX for samples from the ARAB province (Figs. 6 and 7). However, seasonal oscillations in atmospheric forcing of the region, and intermittent complex mesoscale circulation patterns characterizing this basin during all seasons

(Shi et al., 1999), make it a known highly dynamic region for rapid phytoplankton successions, with multiple taxa co-existing during a given season (Garrison et al., 1998; Shalapyonok et al., 2001; Roy et al., 2015). This is supported by the numerous instances of mixed assemblages (“co-existence”, green asterisks in Fig. 7) observed in this region. Typically, the influx of nutrients during boreal summer in the ARAB province as the southwest monsoon deepens the mixed layer leads to a classical diatom-*Phaeocystis* succession pattern nearshore as the nutrient silicate is depleted (Garrison et al., 1998). Contrary to this, we observe an abundance of chlorophytes nearshore in the northeast Arabian Sea during summer, with mixed communities of low relative abundances offshore (Figs. 6 and 7), which is actually consistent with HPLC observations by Latasa and Bidigare (1998) and Goericke (2002) for this region. The lack of dominance by diatoms in the community estimated from pigments during JGOFS was attributed to a weaker-than-usual upwelling rate in the particular summer of 1995, resulting in reduced nutrient enrichment of the basin relative to other years; however, the possibility that the JGOFS observations preceded (missed) the diatom bloom was also acknowledged (Latasa and Bidigare, 1998; Goericke, 2002).

The onset of the northeast monsoon in autumn results in a shallower pycnocline and relaxation of upwelling in the Arabian Sea, leading to mixed assemblages as well as dominance by picophytoplankton offshore, in particular *Prochlorococcus* (Goericke, 2002; Roy et al., 2015), thus supporting our observations of co-existence states and CYANO-4 abundance offshore in autumn in the Arabian Sea (Fig. 7). During winter in this region, evaporative cooling of the surface layer by dry winds leads to mixing of surface waters with the nutricline, typically triggering diatom blooms (Rixen et al., 2006), which is the most abundant phytoplankton we observe nearshore in winter in the ARAB province (Fig. 7). Overall, the spatial and seasonal variability we observe is in agreement with previous pigment analyses in this region and indicate that, at the seasonal scale, no ‘average’ climatological community description suffices for the ARAB province.

3.2.3. North Atlantic Provinces

The region of the northeast Atlantic between the Azores and Portugal, where waters span the transition region between the contiguous NASE and NADR provinces, has been well sampled in space and time by the present dataset (Peloquin

et al., 2013). This area is considered to be one of the most productive open ocean regions owing to the frequent cyclonic and anticyclonic eddy fields (Falkowski et al., 1991; Garçon et al., 2001). Observations from the POMME survey (Maixandeau et al., 2005) investigating community production within these mesoscale features indicate that episodic blooms of diatoms complement an annual backdrop of pico- to nano-sized autotrophic populations. This provides an explanation for the several distinct community patterns observed in spring across this region, exhibiting high relative abundances of either DIATOM-1, HAPTO-6 or cyanobacteria, respectively (Fig. 7). We also observe a community abundant with cyanobacteria in the NASE province in summer, which is known to characterize the eastern North Atlantic provinces when chlorophyll levels are at a relative annual low (Li and Harrison, 2001).

3.2.4. Scales of variability in community structure

In general, the high variability in community structure that we observe in provinces with known mesoscale circulation dynamics stands in contrast to the vast expanses of the tropical and subtropical oceans that exhibit minimal variation in community structure across biogeographic provinces (Figs. 6 and 7). Included in this description are several tropical and subtropical Atlantic (NATR, WTRA, ETRA, SATL) and Pacific provinces (NPTG, SPSG) sampled by the present dataset (Fig. 6). The weak seasonality in temperature and stratification characterizing these provinces within both oceans leads to the perennial abundance of both *Prochlorococcus* and *Synechococcus*, with contributions by other pigment types rarely exceeding 15% Tchla. This is inferred from the CYANO-4 and CYANO-2 abundances dominating provinces in cluster 4 (Fig. 6), and is corroborated by flow cytometry and gene sequencing of cyanobacterial genera along the same transects from which the HPLC climatology derive (Zubkov et al., 2000; Piontkovski et al., 2003; Zwirgmaier et al., 2008). Thus, a description of community structure at the basin scale seems sufficient for open ocean picoplanktonic regimes where the predominant circulation patterns, thus the mixed layer, are governed mainly by geostrophy (Bouman et al., 2011).

A similar observation is made in certain high latitude waters. The Atlantic Arctic Province (ARCT, with samples in the northwest Atlantic) and the South Subtropical Convergence Province (SSTC, with samples in the southwestern Atlantic off the Patagonian shelf; Fig. 6), which are situated in different hemispheres and

biomes from each other (Longhurst, 2007), also share highly similar community structure. The comparable dominance of HAPTO-6 in these two provinces (Figs. 6 and 7) are supported by multiple field surveys in the regions indicating these to be sites of highest recorded global coccolithophore abundances (Gibb et al., 2001; Leblanc et al., 2009; O'Brien et al., 2013; Balch et al., 2014). (It is suggested that silicate limits diatom growth beyond the spring bloom, opening a niche of a moderately stratified water column with excess nitrate for coccolithophores to exploit; Leblanc et al., 2009). Such examples of community analogs in regions that are otherwise geographically remote from one another suggest 'teleconnections' in ecological niche space for plankton in the global ocean (Longhurst, 2007; Zwirgmaier et al., 2008; Kavanaugh et al., 2013).

Overall a province description is likely an oversimplification for the seasonal phytoplankton community in regions where the seasonal mixed layer is interrupted by mesoscale horizontal or vertical water mass features with variable nutrient and temperature profiles. Eddies, fronts, and transition regions among water masses have been proposed as diversity "hotspots" of phytoplankton community composition (e.g., Franks et al., 1992; Barton et al., 2010; d'Ovidio et al., 2010). While we can not rule out the possibility that the sub-province (or sub-seasonal) phytoplankton diversity within the Antarctic, Arabian Sea, and North Atlantic provinces stands out due to the high frequency of samples in these interest regions, which other regions in the dataset are lacking, forecasting the dynamics and changes in such regimes clearly requires a higher degree of complexity in models than is required of ecosystems inhabiting more steady-state hydrographic conditions such as the oligotrophic gyres.

3.3 Drivers of global plankton community structure

In an effort to quantitatively explore the associations of the individual pigment types with environmental conditions, we conducted a simple bivariate analysis between each pigment type and properties of the mixed layer known to influence phytoplankton growth. The surface wind stress magnitude (τ ; N m^{-2}), Ekman upwelling velocity (ω ; m d^{-1}) and depth of the mixed layer (MLD) indirectly affect phytoplankton through vertical mixing and stratification, while the mixed layer median light level (I_{MLD} ; $\text{Ein m}^{-2} \text{d}^{-1}$), mean nitrate concentrations (N_{MLD} ; μM) and mean temperature (T_{MLD} ; $^{\circ}\text{C}$) directly affect phytoplankton growth. Mean total chlorophyll *a* in the mixed layer (CHL_{MLD} ; mg m^{-3} ; calculated from the present

dataset) is included as a trophic proxy to assess its associations with each pigment type. All environmental properties are invariably statistically related to one another (Table 5).

	τ	ω	MLD	Chl _{MLD}	I _{MLD}	N _{MLD}	T _{MLD}
τ	1	0.22	0.55	-0.10	-0.73	0.68	-0.77
ω	-	1	0.11	0.09	-0.17	0.17	-0.33
MLD	-	-	1	-0.10	-0.54	0.38	-0.18
Chl _{MLD}	-	-	-	1	-0.12	0.08	-0.27
I _{MLD}	-	-	-	-	1	-0.61	0.69
N _{MLD}	-	-	-	-	-	1	-0.65
T _{MLD}	-	-	-	-	-	-	1
CYANO-4	-0.44	-0.26	-0.22	-0.26	0.59	-0.56	0.65
CYANO-2	-0.37	-0.13	-0.18	-0.24	0.54	-0.52	0.57
HAPTO-6	0.32	0.09	0.29	ns	-0.25	-0.12	ns
HAPTO-8	0.28	0.24	0.10	-0.23	-0.21	-0.16	-0.18
CHLORO-1	0.10	-0.01	0.07	ns	-0.17	ns	ns
CRYPTO-1	0.26	0.14	0.13	0.25	-0.25	0.29	-0.35
DINO-1	ns	ns	ns	0.05	ns	ns	ns
DIATOM-1	0.18	0.14	ns	0.41	-0.33	0.66	-0.63
CYANO-1	-0.06	ns	ns	ns	0.09	-0.14	0.14

Table 5. Correlation coefficient (r) between environmental properties and relative Tchla of the pigment types. Correlations are significant at the 95% confidence level ($p < 0.05$) unless specified ('ns'). τ = wind stress magnitude ($N\ m^{-2}$), ω = Ekman upwelling velocity ($m\ d^{-1}$), MLD = mixed layer depth (m), CHL_{MLD} = mixed layer mean total chlorophyll a ($ng\ L^{-1}$), I_{MLD} = mixed layer median irradiance ($Ein\ m^{-2}\ d^{-1}$), N_{MLD} = mixed layer mean nitrate concentration (μM), T_{MLD} = mixed layer mean temperature ($^{\circ}C$). Values of r in bold signify the three highest correlated environmental factors for each pigment type (values of $r < 0.10$ were not bolded.)

The strongest linear relationships of CYANO-4 and CYANO-2 are observed with I_{MLD} , N_{MLD} , and T_{MLD} (Table 5), with these properties explaining between 27 – 42% of the variability in relative abundance. CYANO-4 and CYANO-2 are positively correlated with I_{MLD} and T_{MLD} and negatively correlated with N_{MLD} , consistent with the oligotrophic state of waters in which cyanobacteria dominate. Significant negative correlations of τ with CYANO-4 ($r = -0.44$) and CYANO-2 ($r = -0.37$) are observed, as these pigment types are able to exploit highly stratified environments (Bouman et al., 2011). Weak but significant relationships are observed amongst nearly all environmental variables and pigment types, with the exception of DINO-1 and CYANO-1, which exhibit very low abundance (thus low dynamic range in variability) throughout the global dataset. DIATOM-1 exhibits a strong negative association with T_{MLD} ($r = -0.63$) and strong positive association with N_{MLD} ($r = 0.66$), expected as

eutrophic niches favoring diatom growth are typically characterized by high nutrient, cooler upwelled waters (Rixen et al., 2006; Mendes et al., 2013; Alvain et al., 2013). Surprisingly, ω shows rather weak associations ($r = -0.01$ to -0.26) with the main pigment types, possibly because data within $3^{\circ}\text{S} - 3^{\circ}\text{N}$ (where wind stress curl is undefined), and thus estimates within the Pacific equatorial upwelling, were excluded from the regression.

With respect to the nanoplankton, HAPTO-6 and HAPTO-8 have the strongest positive correlation with wind stress magnitude ($r = 0.32$ and $r = 0.28$, respectively). We interpret this as consistent with observations that haptophytes exhibit high abundances in well-mixed waters (Arrigo et al., 1998; Wright and van den Eenden, 2000; Bouman et al., 2011). The cosmopolitan distribution (Fig. 8) and known phenotypic plasticity of coccolithophores and *Phaeocystis* (Vaulot et al., 1994; Schoemann et al., 2005; Zondervan, 2007; Vogt et al., 2012; O'Brien et al., 2013) likely explains why these pigment types exhibit significant but weak correlations with light, nutrients and temperature, as they are able to exploit a wide range of global habitats.

Patterns in community composition likely respond nonlinearly to changes in the environment (Moore et al., 2002; Boyd et al., 2010), particularly where rapid temporal succession of phytoplankton occurs. Based on our earlier observations of variability within dynamic Longhurst provinces (e.g., NASE, NABE, ARAB), it is expected that no universal variable or combination of variables afford a full explanation for relative abundances of different phytoplankton. Nevertheless, in addition to confirming the relative importance of several environmental factors to abundance of pigment types, the linear correlations shown serve as an additional validation measure to our overall CHEMTAX estimates, as up to 42% of variation in relative phytoplankton abundance can be attributed to bottom-up controls.

3.4 Latitudinal gradients in phytoplankton distribution at the global scale

To first order, expected meridional gradients in relative phytoplankton abundance and the extent of each pigment type is readily observed at the global scale and across all seasons (Fig. 8). For example, *Synechococcus* (CYANO-2) exhibits a wider latitudinal range than its picoplankton counterpart, *Prochlorococcus* (CYANO-4), but averages roughly half the abundance of *Prochlorococcus* in the tropics and subtropics where cyanobacteria dominate as a whole, which is consistent with prior

global observations (Partensky et al., 1999; Zwirgmaier et al., 2008; Buitenhuis et al., 2012). The nanoplankton (HAPTO-6, HAPTO-8, CHLORO-1, and CRYPTO-1) exhibit the most cosmopolitan abundance, as microplankton (mainly DIATOM-1 and DINO-1) comprise a very small fraction (<10% on average) equatorward of 40° in both hemispheres, which even includes a number of sampled coastal areas (Fig. 8). It should be noted that due to a high degree of spatial and temporal sampling bias, the seasonal zonal medians of the pigment types are susceptible to over-interpretation as, for example, the high HAPTO-6 abundance at 80°S in summer (Fig. 8) is not only biased to the Ross Sea, but HAPTO-6 estimates in this region are likely all representing *Phaeocystis* (see section 3.2.1).

FIGURE 8

3.5. Global estimates of pigment type-specific chlorophyll *a* and carbon biomass

Annual averages for chlorophyll and carbon fractions were computed to provide a global summary of the relative importance of each pigment type in the mixed layer (Fig. 9). Carbon biomass fractions were determined by applying carbon-to-chlorophyll ratios of 50 for microplankton, and 125 for pico- and nanoplankton to the absolute chlorophyll concentrations of the pigment types, following Le Quéré et al. (2005).

FIGURE 9

Approximation of size classes from the CHEMTAX pigment type fractions can be achieved by summing CYANO-4 and CYANO-2 to picoplankton, HAPTO-6, HAPTO-8, CHLORO-1 and CRYPTO-1 to nanoplankton, and DIATOM-1, DINO-1 and CYANO-1 to microplankton (Finkel et al., 2010). Size classes estimated from CHEMTAX thus comprise 26%, 50% and 24% of global chlorophyll biomass for pico-, nano- and microplankton, respectively. Diagnostic pigment analysis (DPA) using weighting factors from Uitz et al. (2006), when applied to the HPLC dataset, yields global averages of size classes in near equal proportion to each other: 31%, 37% and 32% for pico-, nano- and microplankton, respectively. Estimates of size class from abundance-based remote sensing methods (Hirata et al., 2011), which have been validated using DPA on an independent HPLC dataset, give global estimates for

pico-, nano- and microplankton of 44%, 46% and 11%, respectively. Hirata et al. (2011) supply the caveat that picoplankton are likely overestimated in their method at the expense of microplankton due to fundamental undersampling issues affecting the Southern Ocean; however, the remote sensed nanoplankton estimate is closer to nanoplankton estimated from CHEMTAX than that from DPA. While DPA is a useful and valid approximation for many ecological applications, results suggests that nanoplankton are underestimated at the expense of microplankton using DPA, due to over-allocation of the pigment fucoxanthin to microplankton. At the same time, any estimate of nanoplankton based on the pigment hex will be biased as hex is a significant marker for haptophytes, which recent studies indicate are a highly diverse phyla with a substantial fraction of picoeukaryotic algae (Liu et al., 2009). Further, hex is the dominant pigment in dinoflagellates with chloroplasts derived from haptophyte endosymbionts (e.g., DINO-2, Jeffrey et al., 2011), which are often common or even dominant in some marine environments, and vary in size from nano- to microplankton.

The carbon fractions of the size classes converted from CHEMTAX estimates, 29%, 51% and 20% for pico-, nano- and microplankton, respectively (Fig. 9), compare well with those reported by Le Quéré (2005): 36% pico-, 50% nano- and 14% microplankton. Intracellular carbon-to-chlorophyll ratios vary widely in nature and especially over large iron-limited regimes such as the Southern Ocean, which contains a high frequency of samples in our dataset. The growth rate of cells in iron-limited regions is dramatically reduced relative to expected rates given intracellular chlorophyll concentration (Behrenfeld et al., 2013), suggesting a carbon-to-chlorophyll ratio that averages much lower than 50 in these regions. Given the bias toward Southern Ocean samples in diatom abundance in our dataset, this may explain the slight offset between our microplankton carbon estimate and that of Le Quéré et al. (2005).

The combined estimate of pico- and nanoplankton from CHEMTAX compares well with modeled estimates of this fraction (e.g., Jin et al., 2006). Diatoms receive much explicit representation in prognostic biogeochemical models (Moore et al., 2002; Le Quéré et al., 2005; Jin et al., 2006), but are one-fifth or less of total chlorophyll or carbon biomass herein. A key finding emerging from the present analysis is the large fraction of haptophytes (approaching 40% of total biomass) in agreement with recent surveys (Liu et al., 2009; de Vargas et al., 2015). This group

should be a target for future study, as the species makeup and biogeochemical functional role(s) of this large haptophyte component are mostly unknown (Liu et al., 2009).

While on the global scale percentages of some groups are small (e.g., 5% cryptophytes), on the regional scale variability in their contribution may be highly relevant. Along the Antarctic Peninsula, for example, warmer years with greater ice melt favors the selection of cryptophytes over diatoms, resulting in less grazing by krill and more grazing by salps, which dramatically alters higher trophic level interactions (Mendes et al., 2013). This further emphasizes the need to focus on which spatial scales need be captured in order to simulate ecosystem function.

3.6. Caveats and additional uncertainty

Many of the environmental factors impacting intracellular carbon-to-chlorophyll ratios as discussed above also affect accessory pigment-to-chlorophyll ratios (DiTullio et al., 2013; Behrenfeld et al., 2013). Optimized pigment ratios, thus final phytoplankton abundances determined from CHEMTAX, are still just approximations. Therefore, implicit uncertainties in the estimates persist due to natural variability in phytoplankton pigmentation caused by any number of factors (e.g., temperature, salinity, light exposure, nutrient availability, physical mixing, as well as regional variability in species and strains). The fidelity of the fits using CHEMTAX may be influenced to different degrees by such factors, but at present we do not have a way to control for or quantify this effect.

Our aim was to take as systematic an approach as possible to CHEMTAX initialization using a global dataset in order to circumvent subjectivity involved in ratio matrix selection, which the method has been historically susceptible to (Higgins et al., 2011). As a result, the analysis was limited to choice of the same 9 global pigment types to encapsulate the total contribution of phytoplankton. Many other unknown or minor group contributions are implicit in the major pigment type estimates, in part due to symbiotic acquisition by phytoplankton of pigments and chloroplasts (e.g. dinoflagellates engulfing diazotrophs). On the contrary, some implicit taxa are known. For example, contributions by chrysophytes and pelagophytes may be implicit in the fuco-containing pigment type distributions (i.e., DIATOM-1, HAPTO-6 and HAPTO-8; Not et al., 2008). Fuco is also contained in dinoflagellates, thus in some coastal sites our determinations may yield overestimates

of the diatom community present and underestimates of dinoflagellate abundance (Higgins et al., 2011).

In an effort to show the climatological state of community composition in the surface ocean, we recognize that interannual or decadal-scale changes in pigment stocks due to, for example, influences from El Niño years (Aiken et al., 2009; Higgins and Mackey, 2000) may be averaged out. Conversely, it is also conceivable that what is interpreted as seasonal variability or between-cluster differences for data in the same region (e.g., AMT data in clusters 3 and 4) may in fact represent interannual variability to some extent, but this remains difficult to distinguish at present.

Finally, although “pigment type” is a different classification than phytoplankton functional type (PFT) or phylum, the results may be extrapolated (with some cautions) to PFT definitions: e.g., DIATOM-1 as ‘silicifiers’, HAPTO-6 as ‘calcifiers’, ‘HAPTO-8’ as ‘DMS-producers’, CYANO-2 + CYANO-4 as ‘picoplankton’, and the remaining pigment types as a background community of ‘mixed autotrophs’ (Buitenhuis et al., 2013; Le Quéré et al., 2005). However, recent studies indicate that coccolithophore species may not be the largest component of haptophyte biomass (Liu et al., 2009; O’Brien et al. 2013). Further, diazotroph information is considerably lacking herein as CYANO-1 estimates were insignificant on the global scale. Other recently published datasets of global diazotrophic carbon abundance (Luo et al., 2012) and genetic sequencing of plankton populations including haptophytes (de Vargas et al., 2015) might help fill such gaps; however, as is especially the case with carbon, a contemporaneous total living phytoplankton carbon estimate is needed to place specific PFT carbon biomass estimates into context.

The global CHEMTAX estimates will ideally be a boon to exploration of the link between community structure and marine ecosystem functioning, which is missing at present (Ward et al., 2013), but future applications of these results must take into account the implicit diversity of the haptophyte fraction, as well as the various aforementioned caveats and uncertainty.

3.7. Guidelines for improving future global chemotaxonomic estimates

The assumptions and ambiguities currently associated with the use of HPLC data and chemotaxonomic methods such as CHEMTAX could be effectively reduced in the future following a few proposed guidelines. For one, marine HPLC data

collection and field survey design must take into account the pervasive need for validation measurements using alternative identification methods such as light microscopy, flow cytometry or genomics. This is critical even if such data are only of a qualitative variety (i.e., presence/absence), or even if samples must be frozen as suitable methods are being developed, as such data are particularly useful for knowing which taxa to quantify from shared pigments in a water sample. This kind of collocated information would have been a tremendous advantage in the present study.

Secondly, at least for the foreseeable future, researchers should refer to the updated and comprehensive tables of plankton pigment types found in Jeffrey et al. (2011) and Higgins et al. (2011) as a central resource for classification and guidance in HPLC analysis. In addition to the suite of pigments used in this study, HPLC quantification of other valuable biomarker pigments would confer greater specificity (especially for the diverse haptophyte group) as well as the ability to analyze for more than 7 pigment types in sample when using inverse methods. To this end, we recommend analyzing for chlorophylls c2-MGDG (for more sensitive haptophyte distinctions, e.g., *Emiliana huxleyi*), c1, c2 and c3 (for diatom distinctions), and Mg-DVP (for prasinophyte distinctions). Several additional carotenoids, while some shared, lend greater specificity as well: myxoxanthophyll (for *Trichodesmium* spp. distinction), prasinoxanthin and lutein (for distinguishing among prasinophytes and chlorophytes), and gyroxanthin diesters (for dinoflagellate distinction). Continuous measurement of DVChla and DVChlb is also encouraged, as these remain valuable biomarkers for the genus *Prochlorococcus*, considered to be the most abundant photosynthesizing organism on Earth, and whose biogeographic range will likely expand in a future warmer ocean (Flombaum et al., 2013).

Lastly, it goes without say that HPLC data should be subjected to standardized quality control protocol (e.g., NASA's SeaHARRE program; Hooker and van Heukelem, 2011) for consistency and confidence in the global pigment data collective. Of course, we acknowledge that regional scale studies have specific aims and even more specific budgets on which to conduct investigations, but the present analysis demonstrates that HPLC data from any local or temporal scale can become globally relevant when scaled to a baseline climatology that may aid in the validation and ongoing improvement of ecosystem and remote sensing models of biomass.

4. Conclusion

We have presented a global view of the prevailing dominance patterns in the phytoplankton community and discussed community structure at ocean basin and biogeographic province scales. The novelty of this work is that it is the first globally consistent use of HPLC field observations exploiting inverse methods in order to provide a seasonal baseline of autotrophic community structure, and demonstrates that CHEMTAX may be successfully applied at the global scale. Although we recognize that a great deal of variability is likely concealed through the spatial and temporal averaging of data in favor of this, overall the climatology of phytoplankton pigments unveils a valuable whole community perspective at all the locations sampled and draws attention to the large yet relatively understudied haptophyte component of the global algal population. On the whole, our observations of global community structure can be distilled into two broad descriptions: 1) high diversity mixed assemblages and successions that likely arise due to intraseasonal, regional mesoscale dynamics and 2) picoplankton-dominated communities that exploit basin-scale common environmental niche space.

Plankton are sensitive indicators of change, responding nonlinearly to minor perturbation. It is therefore paramount to support efforts toward synoptically detecting specific phytoplankton abundance, such that climate-induced shifts in ecosystems may be tracked. Recent hindcast evaluations as well as future projections describe environmentally driven community shifts affecting the relative proportions of plankton size classes (Polovina and Woodworth, 2012; Laufkoetter et al., 2013) as well as plankton niches and overall dominance patterns (Alvain et al., 2013; Flombaum et al., 2013; Winter et al., 2014). The Arctic Ocean, for example, may be already demonstrating a tendency from polar to temperate conditions in terms of the bloom dynamics of phytoplankton in this region (Ardyna et al., 2014), and linkages between the shift in onset of polar sea ice retreat with that of community composition has been documented in both hemispheres (Fujiwara et al., 2014; Mendes et al., 2013). Whether such changes are in response to natural or anthropogenic climate shifts, they stress the value of having a baseline seasonal climatology of plankton against which to conduct climate perturbation effects on ocean biogeochemistry. As were the aims of the present study, increasing our attention to the composition and complexity of marine ecosystems leads to determining which environmental domains are most crucial to focus on in the future with respect to their susceptibility to change (Le Quéré et al., 2005; Boyd and Doney, 2009; Boyd et al., 2010).

Acknowledgements

The work was made possible by funds from the Swiss National Science Foundation granted to N. Gruber and M. Vogt. The authors gratefully acknowledge the support of the MAREDAT data contributors, Simon Wright of the Australian Antarctic Division who provided initial guidance and the original CHEMTAX code, and two reviewers for their time and constructive consideration of the work.

References

- Aiken, J., Pradhan, Y., Barlow, R., Lavender, S., Poulton, A., Holligan, P., Hardman-Mountford, N., 2009. Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995 – 2005. *Deep-Sea Res. II* 56, 899 – 917.
- Alvain, S., Moulin, C. Dandonneau, Y., Loisel, H., 2008. Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view. *Global Biogeochem. Cy.* 22, GB3001, doi:10.1029/2007GB003154.
- Alvain, S., Le Quéré, C., Bopp, L., Racault, M-F., Beaugrand, G., Dessailly, D., Buitenhuis, E.T., 2013. Rapid climatic driven shifts of diatoms at high latitudes. *Rem. Sens. Environ.* 132, 195 – 201.
- Andersen, R., Bidigare, R., Keller, M., Latasa, M., 1996. A comparison of HPLC signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. *Deep-Sea Res.* 43, 517 – 537.
- Ardyna, M., Babin, M., Gosselin, M., Devred, E., Rainville, L., Tremblay, J.-E., 2014. Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms. *Geophys. Res. Lett.* 41, 6207 – 6212.
- Arrigo, K.R., Robinson, D.H., Worthen, D.L., Dunbar, R.B., DiTullio, G.R., VanWoert, M., Lizotte, M.P., 1998. Phytoplankton community structure and the drawdown of nutrients and CO₂ in the Southern Ocean. *Science* 283, 365 – 367.

Ayata, S.-D., Levy, M., Aumont, O., Sciandra, A., Sainte-Marie, J., Tagliabue, A., Bernard, O., 2013. Phytoplankton growth formulation in marine ecosystem models: Should we take into account photo-acclimation and variable stoichiometry in oligotrophic areas? *J. Marine Sys.* 125, 29 – 40.

Balch, W.M., Drapeau, D.T., Bowler, B.C., Lyczkowski, E.R., Lubelczyk, L.C., Painter, S.C., Poulton, A.J., 2014. Surface biological, chemical, and optical properties of the Patagonian Shelf coccolithophore bloom, the brightest waters of the Great Calcite Belt. *Limnol. Oceanogr.* 59, 1715 – 1732.

Barton, A.D., Dutkiewicz, S., Flier, G., Bragg, J., Follows, M.J., 2010. Patterns of diversity in marine phytoplankton. *Science* 327, 1509 – 1511.

Behrenfeld, M.J., Milligan, A.J., 2013. Photophysiological expression of iron stress in phytoplankton. *Ann. Rev. Mar. Sci.* 5, 217 – 246.

Behrenfeld, M.J., Boss, E., Siegel, D.A., Shea, D.M., 2005. Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochem. Cy.* 19, GB1006, doi: 10.1029/2004GB002299.

Borg, I., Groenen, P., 2005. *Modern Multidimensional Scaling: theory and applications* (2nd ed.), Springer-Verlag, pp. 207 – 212.

Boyd, P. and Doney, S., 2002. Modelling regional responses by marine pelagic ecosystems to global climate change. *Geophys. Res. Lett.* 29, doi:10.1029/2001GL014130.

Boyd, P.W., Strzepek, R., Fu, F., Hutchins, D.A., 2010. Environmental control of open-ocean phytoplankton groups: Now and in the future. *Limnol. Oceanogr.* 55, 1353 – 1376.

Bouman, H.A., Ulloa, O., Barlow, R., Li, W.K.W., Platt, T., Zwirgmaier, K., Scanlan, D.J., Sathyendranath, S., 2011. Water-column stratification governs the

community structure of subtropical marine picophytoplankton. *Environ. Microbiol. Rep.* 3, 473 – 482.

Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., 2009. Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosci.* 6, 751 – 764.

Brewin, R.J., Lavender, S.J., Hardman-Mountford, N.J., Hirata, T., 2010. A spectral response approach for detecting dominant phytoplankton size class from satellite remote sensing. *Acta Oceanol. Sin.* 29, 14 – 32.

Buitenhuis, E.T., Li, W.K.W., Vaultot, D., Lomas, M.W., Landry, M.R., Partensky, F., Karl, D.M., Ulloa, O., Campbell, L., Jacquet, S., Lantoine, F., Chavez, F., Macias, D., Gosselin, M., McManus, G.B., 2012. Picophytoplankton biomass distribution in the global ocean. *Earth Syst. Sci. Data* 4, 37 – 46, doi:10.5194/essd-4-37-2012.

Buitenhuis, E. T., Vogt, M., Moriarty, R., Bednarsek, N., Doney, S. C., Leblanc, K., Le Quéré, C., Luo, Y.-W., O'Brien, C., O'Brien, T., Peloquin, J., Schiebel, R., Swan, C., 2013. MAREDAT: towards a World Ocean Atlas of MARine Ecosystem DATA. *Earth Syst. Sci. Data* 5, 1077 – 1106, doi:10.5194/essdd-5-1077-2012.

Campbell, L., Nolla, H.A., Vaultot, D., 1994. The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. *Limnol. Oceanogr.* 39, 954 – 961.

Capone, D.G., Zehr, J.P., Paerl, H.W., Bergman, B., Carpenter, E.J., 1997. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* 276, 1221 – 1229.

Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B., Welschmeyer, N.A., 1988. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* 334, 340 – 343.

Claustre, H., 1994. Phytoplankton pigment signatures of the trophic status in various ocean regimes. *Limnol. Oceanogr.* 39, 1206 – 1211.

Dandonneau, Y., Niang, A., 2007. Assemblages of phytoplankton pigments along a shipping line through the North Atlantic and tropical Pacific. *Prog. Oceanogr.* 73, 127 – 144.

Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* PAMI-1 2, 224 – 227.

DiTullio, G.R., Geesey, M.E., Leventer, A., Lizotte, M.P., 2013. Algal pigment ratios in the Ross Sea: Implications for CHEMTAX analysis of Southern Ocean data, In: *Biogeochemistry of the Ross Sea, Antarctic Research Series, Vol. 78*, pp. 35 – 52.

Doney, S.C., Lima, I., Moore, J.K., Lindsay, K., Behrenfeld, M.J., Westberry, T.K., Mahowald, N., Glover, D.M., Takahashi, T., 2009. Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. *J. Mar. Sys.* 76, 95 – 112.

Falkowski, P., 2012. Ocean Science: The power of plankton. *Nature* 483, S17 – S20.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237 – 240.

Finkel, Z.V., Beardall, J., Flynn, K.J., Quigg, A., Rees, T.A.V., Raven, J., 2010. Phytoplankton in a changing world: cell size and elemental stoichiometry. *J. Plank. Res.* 32, 119 – 137.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincon, J., Zabala, L.L., Jiao, N., Karl, D.M., Li, W.K.W., Lomas, M.W., Veneziano, D., Vera, C.S., Vrugt, J.A., Martiny, A.C., 2013. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl Acad. Sci.* 110, 9824 – 9829.

Franks, P.J.S., 1992. Sink or swim: accumulation of biomass at fronts. *Mar. Ecol. Prog. Ser.* 82, 1 – 12.

Fujiwara A., Hirawake, T., Suzuki, K., Imai, I., Saitoh, S.-I., 2014. Timing of sea ice retreat can alter phytoplankton community structure in the western Arctic Ocean. *Biogeosciences* 11, 1705 – 1716.

Garçon, V.C., Oschlies, A., Doney, S.C., McGillicuddy, D., Waniek, J., 2001. The role of mesoscale variability on plankton dynamics in the North Atlantic. *Deep-sea Res. II* 48, 2199 – 2226.

Garrison, D.L., Gowing, M.M., Hughes, M.P., 1998. Nano- and microplankton in the northern Arabian Sea during the Southwest Monsoon, August – September 1995: A US-JGOFS study. *Deep-Sea Res. II* 45, 2269 – 2299.

Gibb, S.W., Cummings, D.G., Irigoien, X., Barlow, R.G., Fauzi, R., Mantoura, C., 2001. Phytoplankton pigment chemotaxonomy of the northeastern Atlantic. *Deep-Sea Res. II* 48, 795 – 823.

Goericke, R., Montoya, J., 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.

Goericke, R., 2002. Top-down control of phytoplankton biomass and community structure in the monsoonal Arabian Sea. *Limnol. Oceanogr.* 47, 1307 – 1323.

Gravalosa, J.M., Flores, J.A., Sierro, F.J., Gersonde, R., 2008. Sea surface distribution of coccolithophores in the eastern Pacific sector of the Southern Ocean (Bellingshausen and Amundsen Seas) during the late austral summer of 2001. *Mar. Micropaleontol* 69, 16 – 25.

Higgins, H.W., Mackey, D.J., 2000. Algal class abundances, estimated from chlorophyll and carotenoid pigments, in the western Equatorial Pacific under El Niño and non-El Niño conditions. *Deep-Sea Res. I* 47, 1461 – 1483.

Higgins, H., Wright, S., Schlüter, L., 2011. Quantitative interpretation of chemotaxonomic pigment data, in: *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*, edited by: Roy, S., Llewellyn, C., Egeland, E., and Johnsen, G., Cambridge University Press, pp. 257 – 313.

Hirata, T., Hardman-Mountford, N., Brewin, R., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., Yamanaka, Y., 2011. Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosci. Disc.* 8, 311 – 327.

Jeffrey, S.W., Wright, S.W., Zapata, M., 2011. Microalgal classes and their signature pigments, in: *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*, edited by: Roy, S., Llewellyn, C., Egeland, E., and Johnsen, G., Cambridge University Press.

Jouini, M., Levy, M., Crepon, M., Thiria, S., 2013. Reconstruction of satellite chlorophyll images under heavy cloud coverage using a neural classification method. *Remote Sens. Env.* 131, 232 – 246.

Kavanaugh, M. T., Hales, B. Saraceno, M., Spitz, Y. H., White, A. E., Letelier, R. M., 2013. Hierarchical and dynamic seascapes: A quantitative framework for scaling pelagic biogeochemistry and ecology. *Prog Oceanogr.* 120, 291 – 304.

Kostadinov, T., Siegel, D., Maritorena, S., 2010. Global variability of phytoplankton functional types from space: assessment via the particle size distribution. *Biogeosci.* 7, 3239 – 3257.

Kozłowski, W.A., Deutschmann, D., Garibotti, I., Trees, C., Vernet, M., 2011. An evaluation of the application of CHEMTAX to Antarctic coastal pigment data. *Deep-Sea Res. I* 58, 350 – 364.

Latasa, M., 2007. Improving estimations of phytoplankton class abundances using CHEMTAX. *Mar. Ecol. Prog. Ser.* 329, 13 – 21.

Latasa, M., Bidigare, R., 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. II* 45, 2133 – 2170.

Laufkoetter, C., Vogt, M., Gruber, N., 2013. Long-term trends in ocean plankton production and particle export between 1960 – 2006. *Biogeosci.* 10, 7373 – 7393.

Leblanc, K., Hare, C.E., Feng, Y., Berg, G.M., DiTullio, G.R., Neeley, A., Benner, I., Sprengel, C., Beck, A., Sanudo-Wilhemly, S.A., Passow, U., Klinck, K., Wilhelm, S.W., Brown, C.W., Hutchins, D.A., 2009. Distribution of calcifying and silicifying phytoplankton in relation to environmental and biogeochemical parameters during the late stages of the 2005 North East Atlantic Spring Bloom. *Biogeosci.* 6, 1 – 25.

Le Quéré, C., Harrison, S., Prentice, I.C., Buitenhuis, E.T., Aumont, O., Bopp, L., Claustre, H., Cotrim da Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfeld, K.E., Legendre, L., Manizza, M., Platt, T., Rivkin, R., Sathyendranath, S., Uitz, J., Watson, A.J., Wolf-Gladrow, D., 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Glob. Change Biol.* 11, 2016 – 2040.

Letelier, R., Bidigare, R., Hebel, D., Ondrusek, M., Winn, C., Karl, D., 1993. Temporal variability of phytoplankton community structure based on pigment analysis. *Limnol. Oceanogr.* 38, 1420 – 1437.

Li, W.K.W., Harrison, W.G., 2001. Chlorophyll, bacteria and picophytoplankton in ecological provinces of the North Atlantic. *Deep-Sea Res. II* 48, 2271 – 2293.

Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F., de Vargas, C., 2009. Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in the open oceans. *Proc. Nat. Acad. Sci.* 106, 12803 – 12808.

Llewellyn, C.A., Fishwick, J., Blackford, J.C., 2005. Phytoplankton community assemblage in the English Channel: a comparison using chlorophyll *a* derived from

HPLC-CHEMTAX and carbon derived from microscopy cell counts. *J. Plankton Res.* 27, 103 – 119.

Longhurst, A., 2007. *Ecology Geography of the Sea*, 2nd Edition. Elsevier Academic Press.

Luo, Y.-W., Doney, S.C., Anderson, L.A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Bostrom, K.H., Bottjer, D., Capone, D.G., Carpenter, E.J., Chen, Y. L., Church, M.J., Dore, J.E., Falcon, L.I., Fernandez, A., Foster, R.A., Furuya, K., Gómez, F., Gundersen, K., Hynes, A.M., Karl, D.M., Kitajima, S., Langlois, R.J., LaRoche, J., Letelier, R.M., Maranon, E., McGillicuddy Jr., D.J., Moisander, P.H., Moore, C.M., Mourino-Carballido, B., Mulholland, M.R., Needoba, J.A., Orcutt, K. M., Poulton, A.J., Rahav, E., Raimbault, P., Rees, A.P., Riemann, L., Shiozaki, T., Subramaniam, A., Tyrrell, T., Turk-Kubo, K.A., Varela, M., Villareal, T.A., Webb, E. A., White, A.E., Wu, J., Zehr, J.P., 2012. Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth Syst. Sci. Data*, 4, 47 – 73, doi:10.5194/essd-4-47-2012.

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, *Mar. Ecol. Prog. Ser.* 144, 265 – 283.

Maixandeau, A., Lefevre, D., Karayanni, H., Christaki, U., Van Wambeke, F., Thyssen, M., Denis, M., Fernandez, C.I., Uitz, J., Leblanc, K., Quequiner, B., 2005. Microbial community production, respiration, and structure of the microbial food web of an ecosystem in the northeastern Atlantic Ocean. *J. Geophys. Res.* 110, C07S17, doi:10.1029/2004JC002694.

Mendes, C.R.B., Tavano, V.M., Leal, M.C., de Souza, M.S., Brotas, V., Garcia, C.A.E., 2013. Shifts in the dominance between diatoms and cryptophytes during three late summers in the Bransfield Strait (Antarctic Peninsula). *Polar Biol.* 36, 537 – 547.

Moore, J.K., Doney, S.C., Kleypas, J.A., Glover, D.M., Fung, I.Y., 2002. An intermediate complexity marine ecosystem model for the global domain. *Deep-Sea Res.*, II 49, 403 – 462.

Mustapha, Z.B., Alvain, S., Jamet, C., Loisel, H., Dessailly, D., 2014. Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery: Application to the detection of phytoplankton groups in open ocean waters. *Rem. Sens. Environ.* 146, 97 – 112.

Not, F., Latasa, M., Scharek, R., Viprey, M., Karleskind, P., Balague, V., Ontoria-Oviedo, I., Cumino, A., Goetze, E., Vaultot, D., Massana, R., 2008. Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeucaryotes. *Deep-Sea Res. I* 55, 1456 – 1473.

O'Brien, C. J., Peloquin, J. A., Vogt, M., Heinle, M., Gruber, N., Ajani, P., Andrulleit, H., Aristegui, J., Beaufort, L., Estrada, M., Karentz, D., Kopczynska, E., Lee, R., Poulton, A. J., Pritchard, T., Widdicombe, C., 2013. Global marine plankton functional type biomass distributions: coccolithophores. *Earth Syst. Sci. Data*, 5, 259 – 276, doi:10.5194/essd-5-259-2013.

d'Ovidio, F., De Monte, S., Alvain, S., Dandonneau, Y., Levy, M., 2010. Fluid dynamical niches of phytoplankton types. *Proc. Natl. Acad. Sci. USA*, 107, 18366 – 18370.

Partensky, F., Blanchot, J., Vaultot, D., 1999. Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: A review, in: *Marine Cyanobacteria*, edited by: Charpy, L., Larkum, A.W.D., Bull. Inst. Oceanogr. Monaco. Special issue 19, pp. 457 – 475.

Peloquin, J., Swan, C., Gruber, N., Vogt, M., Claustre, H., Ras, J., Uitz, J., Barlow, R., Behrenfeld, M., Bidigare, R., Dierssen, H., Ditullio, G., Fernandez, E., Gallienne, C., Gibb, S., Goericke, R., Harding, L., Head, E., Holligan, P., Hooker, S., Karl, D., Landry, M., Letelier, R., Llewellyn, C. A., Lomas, M., Lucas, M., Man- nino, A., Marty, J.-C., Mitchell, B. G., Muller-Karger, F., Nelson, N., O'Brien, C., Prezelin, B.,

Repeta, D., Jr. Smith, W. O., Smythe-Wright, D., Stumpf, R., Subramaniam, A., Suzuki, K., Trees, C., Vernet, M., Wasmund, N., Wright, S., 2013. The MAREDAT global database of high performance liquid chromatography marine pigment measurements. *Earth Syst. Sci. Data* 5, 109 – 123, doi:10.5194/essd-5-109-2013.

Piontkovski, S.A., Landry, M.R., Finenko, Z.Z., Kovalev, A.V., Williams, R., Gallienne, C.P., Mishonov, A.V., Skryabin, V.A., Tokarev, Y.N., Nikolsky, V.N., 2003. Plankton communities of the South Atlantic anticyclonic gyre. *Oceanol. Acta* 26, 255 – 268.

Polovina, J.J., Woodworth, P.A., 2012. Declines in phytoplankton cell size in the subtropical oceans estimated from satellite remotely-sensed temperature and chlorophyll, 1998–2007. *Deep-Sea Res. II* 77-80, 82 – 88.

Raitsos, D., Lavendar, S., Maravelias, C., Haralabous, J., Richardson, A., Reid, P., 2008. Identifying four phytoplankton functional types from space: An ecological approach. *Limonol. Oceanogr.*, 53, 605 – 613.

Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., Maury, O., 2013. Dynamic biogeochemical provinces in the global ocean. *Global Biogeochem. Cy.* 27, 1046 – 1058.

Risien, C.M., Chelton, D.B., 2008. A global climatology of surface wind and wind stress fields from eight years of QuikSCAT scatterometer data. *J. Phys. Oceanogr.* 38, 2379 – 2413.

Rixen, T., Goyet, C., Ittekkot, V., 2006. Diatoms and their influence on the biologically mediated uptake of atmospheric CO₂ in the Arabian Sea upwelling system. *Biogeosci.* 3, 1 – 13.

Roy, R., Chitari, R., Kulkarni, V., Krishna, M.S., Sarma, V.V.S.S., Anil, A.C. CHEMTAX-derived phytoplankton community structure associated with temperature fronts in the northeastern Arabian Sea, 2015. *J. Marine Sys.* 144, 81 – 91.

Schlüter, L., Mohlenberg, F., 2003. Detecting presence of phytoplankton groups with non-specific pigment signatures. *J. Appl. Phycol.* 15, 465 – 476.

Schlüter, L., Henriksen, P., Nielsen, T.G., Jakobsen, H.H., 2011. Phytoplankton composition and biomass across the southern Indian Ocean. *Deep-Sea Res. I* 58, 546 – 556.

Schoemann, V., Becquevort, S., Stefels, J., Rousseau, V., Lancelot, C., 2005. *Phaeocystis* blooms in the global ocean and their controlling mechanisms: a review. *J. Sea Res.* 53, 43 – 66.

Shalapyonok, A., Olson, R.J., Shalapyonok, L.S., 2001. Arabian Sea phytoplankton during Southwest and Northeast Monsoons 1995: composition, size structure and biomass from individual cell properties measured by flow cytometry. *Deep-Sea Res. I* 48, 1231 – 1261.

Sherrard, N.J., Nimmo, M., Llewellyn, C.A., 2006. Combining HPLC pigment markers and similarity indices to assess phytoplankton community structure: An environmental tool for eutrophication? *Sci. Tot. Environ.* 361, 97 – 110.

Shi, W., Morrison, J.M., Bohm, E., Manghnani, V., 1999. Remotely sensed features in the US JGOFS Arabian Sea Process Study. *Deep-Sea Res. II* 46, 1551 – 1575.

Smayda, T., 2002. Adaptive ecology, growth strategies and the global bloom expansion of dinoflagellates. *J. Oceanogr.* 58, 281 – 294.

Smith, R.C., Baker, K.S., Byers, M.L., Stammerjohn, S.E., 1998. Primary productivity of the Palmer Long Term Ecological Research Area and the Southern Ocean. *J. Mar. Sys.* 17, 245 – 259.

Smith Jr., W.O., Dinniman, M.S., Tozzi, S., DiTullio, G.R., Mangoni, O., Modigh, M., Saggiomo, V., 2010. Phytoplankton photosynthetic pigments in the Ross Sea: Patterns and relationships among functional groups. *J. Mar. Sys.* 82, 177 – 185.

- Subramaniam, A., Brown, C.W., Hood, R.R., Carpenter, E.J., Capone, D.G., 2002. Detecting *Trichodesmium* blooms in SeaWiFS imagery. *Deep-Sea Res. II*, 49, 107 – 121.
- Torrecilla, E., Stramski, D., Reynolds, R., Millan-Nunez, E., Piera, J., 2011. Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages. *Remote Sens. Env.* 115, 2578 – 2593.
- Tyrrell, T., Maranon, E., Poulton, A.J., Bowie, A.R., Harbour, D.S., Woodward, E.M.S., 2003. Large-scale latitudinal distribution of *Trichodesmium* spp. in the Atlantic Ocean. *J. Plankton Res.* 25, 405–416.
- Uitz, J., Claustre, H., Morel, A., Hooker, S., 2006. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *J. Geophys. Res.* 111, doi:10.1029/2005JC003207.
- Van den Meersche, K., Soetaert, K., Middelburg, J., 2008. A Bayesian compositional estimator for microbial taxonomy based on biomarker. *Limnol. Oceanogr.: Methods* 6, 190 – 199.
- Van Heukelem, L. and Hooker, S.B., 2011. The importance of a quality assurance plan for method validation and minimizing uncertainties in the HPLC analysis of phytoplankton pigments, in: *Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography*, edited by: Roy, S., Llewellyn, C., Egeland, E., and Johnsen, G., Cambridge University Press, pp. 195 – 242.
- Van Leeuwe, M.A. and Stefels, J., 2007. Photosynthetic responses in *Phaeocystis Antarctica* towards varying light and iron conditions. *Biogeochem.* 83, 61 – 70.
- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., et al., 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348. doi:10.1126/science.1261605.

Vaulot, D., Birrien, J.-L., Marie, D., Casotti, R., 1994. Morphology, ploidy, pigment composition, and genome size of cultured strains of *Phaeocystis* (Prymnesiophyceae). *J. Phycol.* 30, 1022 – 1035.

Vidussi, F., Claustre, H., Bustillos-Guzman, J., Cailliau, C., Marty, J.C., 1996. Determination of chlorophylls and carotenoids of marine phytoplankton: separation of chlorophyll a from divinyl chlorophyll a and zeaxanthin from lutein. *J. Plankton Res.* 18, 2377 – 2382.

Vidussi, F., Claustre, H., Manca, B., Luchetta, A. Marty. J-C., 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *J. Geophys. Res.* 106, 19, 939 – 956.

Vogt, M., O'Brien, C., Peloquin, J., Schoemann, V., Breton, E., Estrada, M., Gibson, J., Karentz, D., Van Leeuwe, M. A., Stefels, J., Widdicombe, C., Peperzak, L., 2012. Global marine plankton functional type biomass distributions: *Phaeocystis* spp. *Earth Syst. Sci. Data* 4, 107 – 120, doi:10.5194/essd-4-107-2012.

Ward, B. A., Schartau, M., Oschlies, A., Martin, A.P., Follows, M.J., Anderson, T.R., 2013. When is a biogeochemical model too complex? Objective model reduction and selection for North Atlantic time-series sites. *Prog. Oceanogr.* 116, 49 – 65.

Winter, A., Henderiks, J., Beaufort, L., Rickaby, R.E.M., Brown, C.W., 2014. Poleward expansion of the coccolithophore *Emiliania huxleyi*. *J. Plankton Res.* 36, 316 – 325.

Wolf, C., Frickenhaus, S., Kiliyas, E.S., Peeken, I., Metfies, K., 2014. Protist community composition in the Pacific sector of the Southern Ocean during austral summer 2010. *Polar Biol.* 37, 375 – 389.

Wright, S.W. and Jeffrey, S.W., 2006. Pigment markers for phytoplankton production, in: *Marine Organic Matter: Biomarkers, Isotopes and DNA*, edited by J.K. Volkman. *The Handbook of Environmental Chemistry* vol. 2, Part N. Springer, Berlin, pp. 71 – 104.

Wright, S.W., van den Enden, R.L., 2000. Phytoplankton community structure and stocks in the East Antarctic marginal ice zone (BROKE survey, January – March 1996) determined by CHEMTAX analysis of HPLC pigment signatures. *Deep-Sea Res. I* 47, 2363 – 2400.

Wright, S.W., Ishikawa, A., Marchant, H.J., Davidson, A.T., van den Enden, R.L., Nash, G.V., 2009. Composition and significance of picophytoplankton in Antarctic waters. *Polar Biol.* 32, 797 – 808.

Wright, S., van den Enden, R., Pearce, I., Davidson, A., Scott, F., Westwood, K., 2010. Phytoplankton community structure and stocks in the Southern Ocean (30 – 80°E) determined by CHEMTAX analysis of HPLC pigment signatures. *Deep-Sea Res. II* 57, 758 – 778.

Yao, P., Yu, Z., Deng, C., Liu, S., Zhen, Y., 2011. Classification of marine diatoms using pigment ratio suites. *Chinese J. Oceanol. Limnol.* 29, 1075 – 1085.

Zapata, M., Jeffrey, S.W., Wright, S.W., Rodriguez, F., Garrido, J.L., Clementson, L., 2004. Photosynthetic pigments in 37 species (65 strains) of Haptophyta: implications for oceanography and chemotaxonomy. *Mar. Ecol. Prog. Ser.* 270, 83 – 102.

Zondervan, I., 2007. The effects of light, macronutrients, trace metals and CO₂ on the production of calcium carbonate and organic carbon in coccolithophores—A review. *Deep-Sea Res. II* 54, 521 – 537.

Zubkov, M.V., Sleigh, M.A., Burkill, P.H., Leakey, R.J.G., 2000. Picoplankton community structure on the Atlantic Meridional Transect: a comparison between seasons. *Prog. Oceanogr.* 45, 369 - 386.

Zwirgmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaultot, D., Not, F., Massana, R., Ulloa, O., Scanlan, D.J., 2008. Global phylogeography of

marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ. Microb.* 10, 147 – 161.

FIGURE CAPTIONS:

Fig. 1. Seasonal composite maps showing the global locations of HPLC pigment data used in the present study. Colored markers indicate to which cluster data belong.

Fig. 2. Dendrogram resulting from hierarchical clustering of the pigment ratios. (Cluster name designations are approximate and were assigned on the basis of the spatial distribution of the data observed in Fig. 1.) *Inset*: Davies-Bouldin index versus number of clusters.

Fig. 3. Median bulk pigment:Tchl_a ratios characterizing each data cluster.

Fig. 4. Annual median community of each cluster represented as a function of a) relative Tchl_a (expressed as % of total) and b) absolute Tchl_a (mg/m³).

Fig. 5. Non-metric multidimensional scaling ordination plot of the data based on their underlying community structure. Colors and markers correspond to which original cluster the data belong. Distance between data points corresponds to degree of similarity/dissimilarity with respect to 9-dimensional pigment type space.

Fig. 6. Seasonal median phytoplankton communities for each Longhurst province contained in clusters 4, 5 and 6, respectively. Pigment types expressed as % Tchl_a. Province codes reproduced from Longhurst (2007). The median community for each displayed province was computed from a minimum of 10 samples.

Fig. 7. Global seasonal composite climatologies of the most abundant phytoplankton pigment type in the mixed layer. The designation of ‘most abundant’ was assigned to the pigment type with the highest relative Tchl_a in excess of 30% (colored circles). Where none or more than one pigment type comprised $\geq 30\%$ of Tchl_a, or where two

or more pigment types made equal contributions $\geq 30\%$ Tchl a , a state of co-existence was assigned (green asterisk).

Fig. 8. Seasonal composite climatologies of the zonal mixed layer median % Tchl a attributed to each phytoplankton pigment type.

Fig. 9. Global averages from the present study of: a) % Tchl a , and b) % total carbon biomass attributed to each phytoplankton pigment type. Carbon biomass fractions were calculated using conversion factors from Le Quéré et al. (2005).

Supplementary Fig. S1. Non-metric multidimensional scaling ordination plot of the data in terms of 7-dimensional pigment ratio space. Colors and markers correspond to which cluster data points belong. Distance between data points corresponds to degree of similarity/dissimilarity with respect to pigment ratios.

Appendix

A.1. Seasonal composites of mixed layer mean % of chlorophyll a per pigment type. This product is 1 x 1 degree relative total chlorophyll a (expressed as fraction of total chlorophyll a) of each phytoplankton pigment type, accessible via: [doi.pangaea.de/10.1594/PANGAEA.855412](https://doi.org/10.1594/PANGAEA.855412)

A.2. Seasonal composites of mixed layer mean total chlorophyll a per pigment type. This product is 1 x 1 degree total chlorophyll a concentration (mg/m^3) of each phytoplankton pigment type, accessible via: [doi.pangaea.de/10.1594/PANGAEA.855412](https://doi.org/10.1594/PANGAEA.855412),

A.3. Initial pigment ratios used in the present study:

	Winter	Spring	Summer	Autumn
Cluster 1	-	-	LL	-
Cluster 2	-	-	LL	-
Cluster 3	-	ML	-	ML
Cluster 4	ML	ML	ML	ML
Cluster 5	LL	LL	LL	LL
Cluster 6	LL	LL	LL	LL

ACCEPTED MANUSCRIPT

Table A3.1. Denotes whether pigment ratios from low-light (“LL”) or medium-light (“ML”) environments were selected from Higgins et al. (2011) for the pigment types evaluated in each season of each cluster.

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-1	0	0	0	0	0	0	0.179	1
CYANO-2	0	0	0	0	0	0	0.358	1
DIATOM-1	0	0	0	0.947	0	0	0	1
HAPTO-6	0	0.016	0	0.224	1.342	0	0	1
HAPTO-8	0	0.688	0	1.032	0.900	0	0	1
DINO-1	0	0	0	0	0	0.838	0	1
CRYPTO-1	0.211	0	0	0	0	0	0	1
CHLORO-1	0	0	0.328	0	0	0	0.027	1

Table A3.2. Initial “LL” pigment ratios. (Final ratio matrices given in A.4.1 – A.4.16 display which pigment types were included for each cluster.)

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-1	0	0	0	0	0	0	0.215	1
CYANO-2	0	0	0	0	0	0	0.498	1
DIATOM-1	0	0	0	0.805	0	0	0	1
HAPTO-6	0	0.015	0	0.186	1.138	0	0	1
HAPTO-8	0	0.655	0	0.577	0.668	0	0	1
DINO-1	0	0	0	0	0	0.851	0	1
CRYPTO-1	0.219	0	0	0	0	0	0	1
CHLORO-1	0	0	0.332	0	0	0	0.043	1

Table A3.3. Initial “ML” pigment ratios. (Final ratio matrices given in A.4.1 – A.4.16 display which pigment types were included for each cluster.)

A.4. Final pigment ratio matrices per cluster per season:

	allox	but	chlb	fuco	hex	peri	zeax	tchla
DIATOM-1	0	0	0	2.554	0	0	0	1
HAPTO-6	0	0.018	0	0.193	1.096	0	0	1
HAPTO-8	0	0.904	0	1.029	1.302	0	0	1
DINO-1	0	0	0	0	0	0.691	0	1
CRYPTO-1	0.291	0	0	0	0	0	0	1
CHLORO-1	0	0	0.145	0	0	0	0.087	1

Table A4.1. Cluster 1 Summer

	allox	but	chlb	fuco	hex	peri	zeax	tchla
DIATOM-1	0	0	0	1.201	0	0	0	1
HAPTO-6	0	0.024	0	0.230	1.380	0	0	1
HAPTO-8	0	0.364	0	0.544	0.697	0	0	1
DINO-1	0	0	0	0	0	0.760	0	1
CRYPTO-1	0.276	0	0	0	0	0	0	1
CHLORO-1	0	0	0.052	0	0	0	0.092	1

Table A4.2. Cluster 2 Summer

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-1	0	0	0	0	0	0	0.226	1

ACCEPTED MANUSCRIPT

CYANO-2	0	0	0	0	0	0	2.822	1
DIATOM-1	0	0	0	0.952	0	0	0	1
HAPTO-6	0	0.056	0	0.083	0.872	0	0	1
DINO-1	0	0	0	0	0	0.782	0	1
CHLORO-1	0	0	0.471	0	0	0	0.057	1

Table A4.3. Cluster 3 Spring

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-1	0	0	0	0	0	0	0.194	1
CYANO-2	0	0	0	0	0	0	2.059	1
DIATOM-1	0	0	0	0.805	0	0	0	1
HAPTO-6	0	0.026	0	0.162	1.281	0	0	1
DINO-1	0	0	0	0	0	0.723	0	1
CHLORO-1	0	0	0.406	0	0	0	0.044	1

Table A4.4. Cluster 3 Autumn

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-1	0	0	0	0	0	0	0.247	1
CYANO-2	0	0	0	0	0	0	0.928	1
DIATOM-1	0	0	0	0.804	0	0	0	1
HAPTO-6	0	0.017	0	0.209	1.207	0	0	1
DINO-1	0	0	0	0	0	0.900	0	1
CHLORO-1	0	0	0.366	0	0	0	0.047	1

Table A4.5. Cluster 4 Winter

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-2	0	0	0	0	0	0	0.949	1
DIATOM-1	0	0	0	0.671	0	0	0	1
HAPTO-6	0	0.017	0	0.061	0.859	0	0	1
HAPTO-8	0	0.571	0	0.094	0.739	0	0	1
DINO-1	0	0	0	0	0	0.763	0	1
CHLORO-1	0	0	0.609	0	0	0	0.074	1

Table A4.6. Cluster 4 Spring

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-2	0	0	0	0	0	0	0.700	1
DIATOM-1	0	0	0	0.571	0	0	0	1
HAPTO-6	0	0.020	0	0.118	1.387	0	0	1
HAPTO-8	0	0.474	0	0.068	0.808	0	0	1
DINO-1	0	0	0	0	0	0.759	0	1
CHLORO-1	0	0	0.561	0	0	0	0.092	1

Table A4.7. Cluster 4 Summer

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-2	0	0	0	0	0	0	0.849	1
DIATOM-1	0	0	0	0.851	0	0	0	1
HAPTO-6	0	0.013	0	0.031	0.815	0	0	1
HAPTO-8	0	0.855	0	0.086	0.953	0	0	1
DINO-1	0	0	0	0	0	0.765	0	1
CHLORO-1	0	0	0.642	0	0	0	0.082	1

Table A4.8. Cluster 4 Autumn

ACCEPTED MANUSCRIPT

	allox	but	chlb	fuco	hex	peri	zeax	tchla
DIATOM-1	0	0	0	0.510	0	0	0	1
HAPTO-6	0	0.014	0	0.231	1.331	0	0	1
HAPTO-8	0	0.804	0	0.594	0.709	0	0	1
DINO-1	0	0	0	0	0	0.798	0	1
CRYPTO-1	0.422	0	0	0	0	0	0	1
CHLORO-1	0	0	0.841	0	0	0	0.287	1

Table A4.9. Cluster 5 Winter

	allox	but	chlb	fuco	hex	peri	zeax	tchla
DIATOM-1	0	0	0	0.619	0	0	0	1
HAPTO-6	0	0.012	0	0.191	1.309	0	0	1
HAPTO-8	0	0.645	0	0.673	0.590	0	0	1
DINO-1	0	0	0	0	0	0.836	0	1
CRYPTO-1	0.541	0	0	0	0	0	0	1
CHLORO-1	0	0	0.298	0	0	0	0.133	1

Table A4.10. Cluster 5 Spring

	allox	but	chlb	fuco	hex	peri	zeax	tchla
DIATOM-1	0	0	0	0.678	0	0	0	1
HAPTO-6	0	0.016	0	0.207	1.173	0	0	1
HAPTO-8	0	0.343	0	0.452	0.483	0	0	1
DINO-1	0	0	0	0	0	0.789	0	1
CRYPTO-1	0.321	0	0	0	0	0	0	1
CHLORO-1	0	0	0.240	0	0	0	0.165	1

Table A4.11. Cluster 5 Summer

	allox	but	chlb	fuco	hex	peri	zeax	tchla
DIATOM-1	0	0	0	0.572	0	0	0	1
HAPTO-6	0	0.014	0	0.243	1.244	0	0	1
HAPTO-8	0	0.783	0	0.880	0.838	0	0	1
DINO-1	0	0	0	0	0	0.900	0	1
CRYPTO-1	0.249	0	0	0	0	0	0	1
CHLORO-1	0	0	0.200	0	0	0	0.026	1

Table A4.12. Cluster 5 Autumn

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-2	0	0	0	0	0	0	0.714	1
DIATOM-1	0	0	0	2.129	0	0	0	1
HAPTO-6	0	0.009	0	0.068	0.474	0	0	1
HAPTO-8	0	0.675	0	0.097	0.788	0	0	1
DINO-1	0	0	0	0	0	0.817	0	1
CRYPTO-1	0.201	0	0	0	0	0	0	1
CHLORO-1	0	0	0.880	0	0	0	0.139	1

Table A4.13. Cluster 6 Winter

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-2	0	0	0	0	0	0	0.854	1
DIATOM-1	0	0	0	1.791	0	0	0	1
HAPTO-6	0	0.010	0	0.067	0.642	0	0	1
HAPTO-8	0	0.630	0	0.219	0.791	0	0	1
DINO-1	0	0	0	0	0	0.935	0	1

ACCEPTED MANUSCRIPT

CRYPTO-1	0.255	0	0	0	0	0	0	1
CHLORO-1	0	0	0.521	0	0	0	0.031	1

Table A4.14. Cluster 6 Spring

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-2	0	0	0	0	0	0	0.939	1
DIATOM-1	0	0	0	1.339	0	0	0	1
HAPTO-6	0	0.012	0	0.053	0.790	0	0	1
HAPTO-8	0	0.414	0	0.063	0.681	0	0	1
DINO-1	0	0	0	0	0	0.832	0	1
CRYPTO-1	0.335	0	0	0	0	0	0	1
CHLORO-1	0	0	0.912	0	0	0	0.091	1

Table A4.15. Cluster 6 Summer

	allox	but	chlb	fuco	hex	peri	zeax	tchla
CYANO-2	0	0	0	0	0	0	0.607	1
DIATOM-1	0	0	0	0.908	0	0	0	1
HAPTO-6	0	0.007	0	0.046	0.444	0	0	1
HAPTO-8	0	1.236	0	0.285	1.783	0	0	1
DINO-1	0	0	0	0	0	0.839	0	1
CRYPTO-1	0.801	0	0	0	0	0	0	1
CHLORO-1	0	0	0.986	0	0	0	0.079	1

Table A4.16. Cluster 6 Autumn

Highlights

First systematic global application of CHEMTAX to HPLC phytoplankton pigments

1500 1°x1° seasonal climatological grid points for 9 phytoplankton groups from HPLC

Haptophytes are a large (>30% total chlorophyll) component of global phytoplankton

Complexity in community structure greatest in regions with known mesoscale dynamics

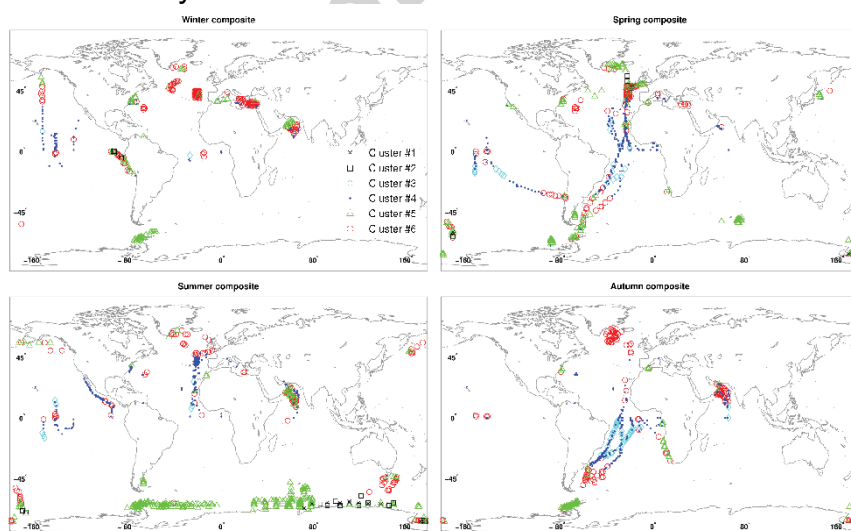


fig 1

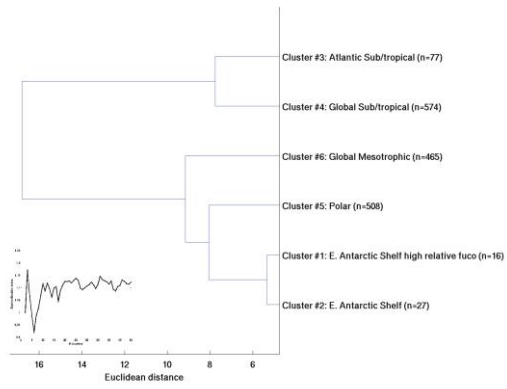


fig 2

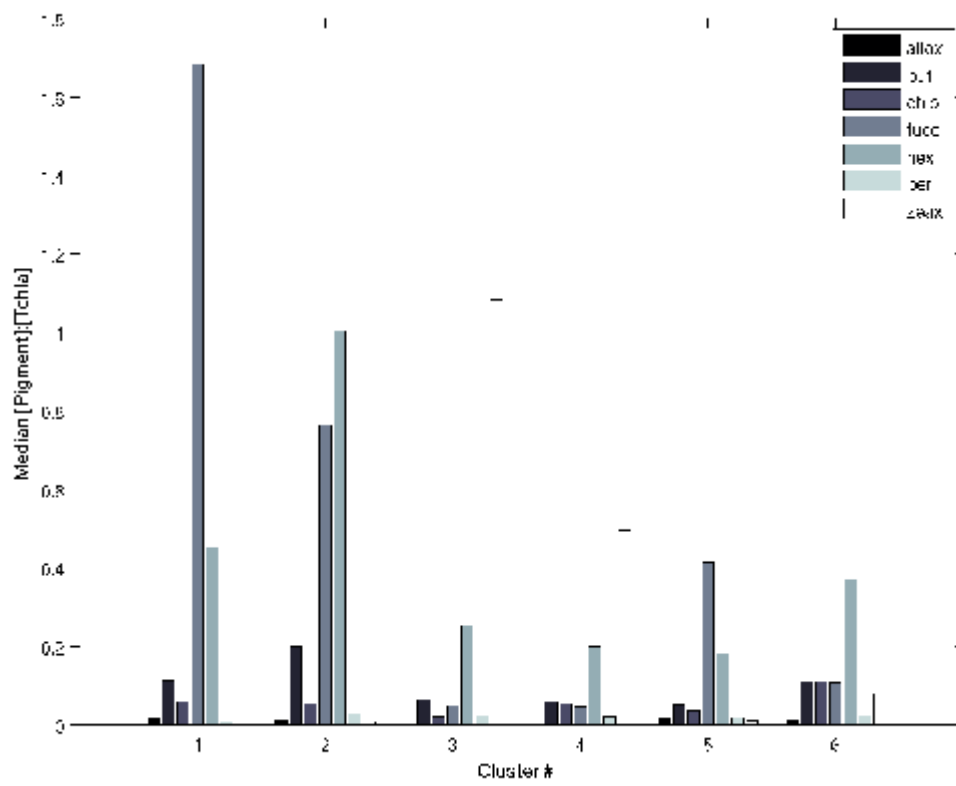


fig 3

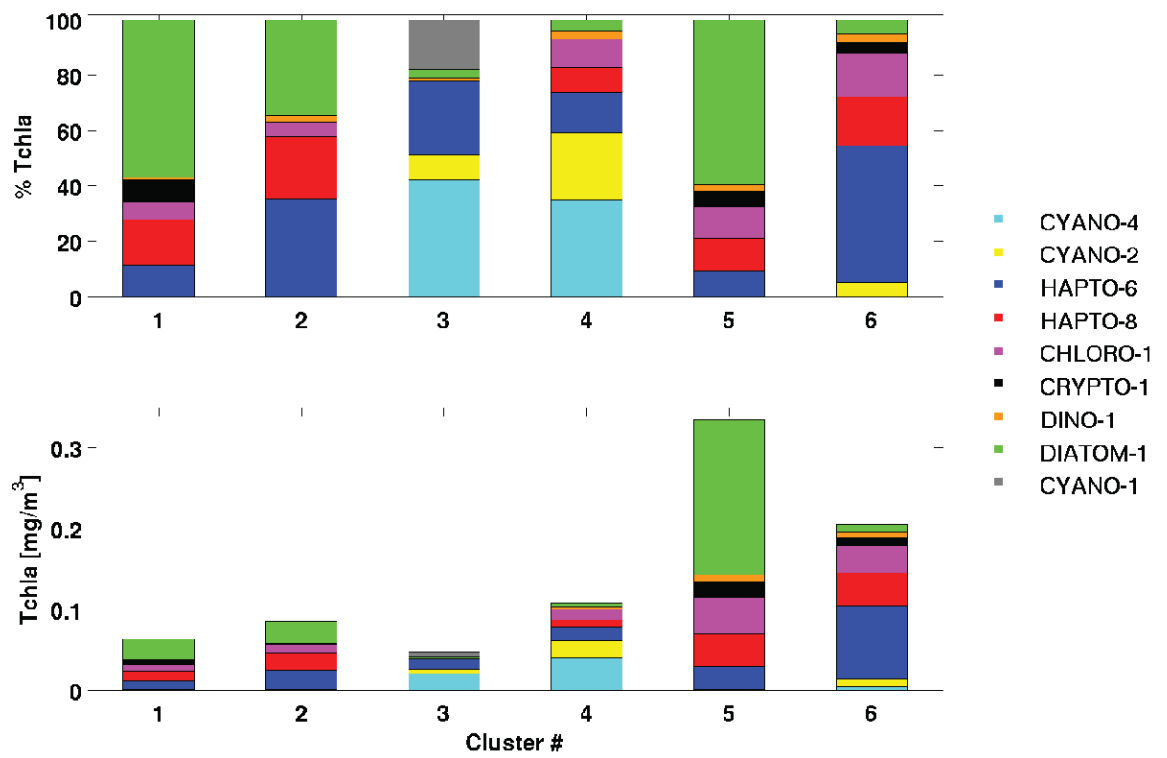


fig 4

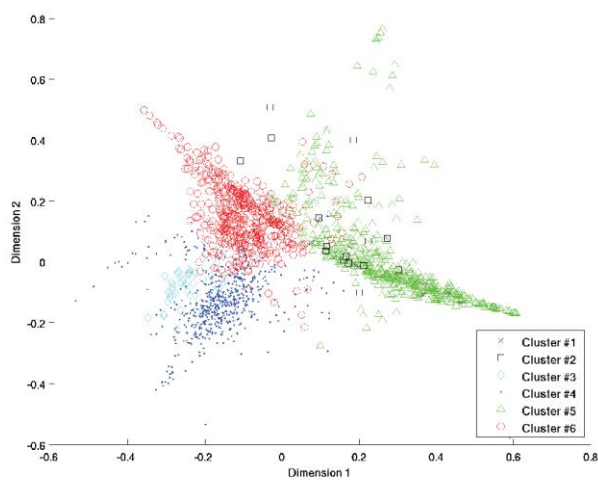


fig 5

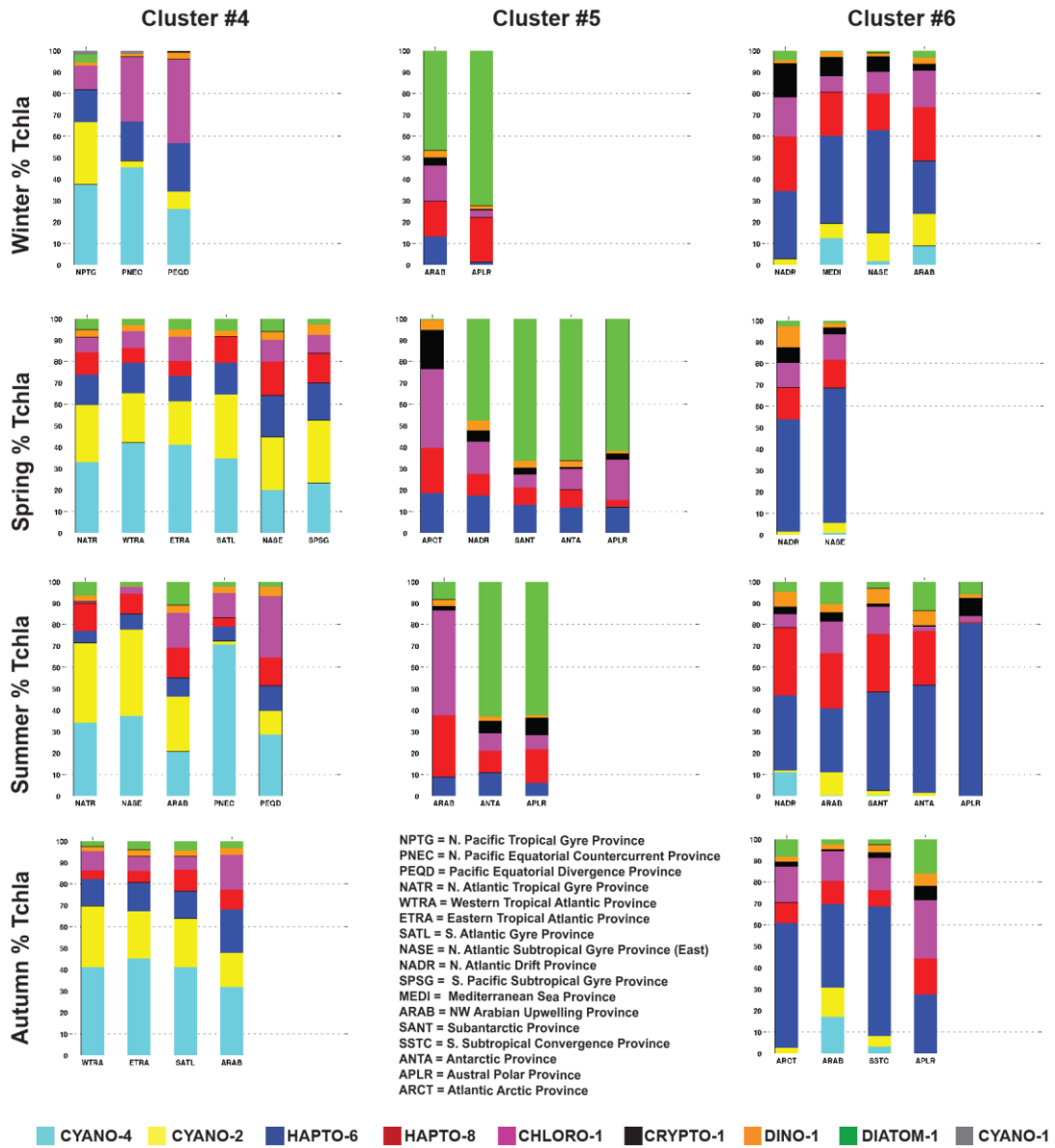


fig 6

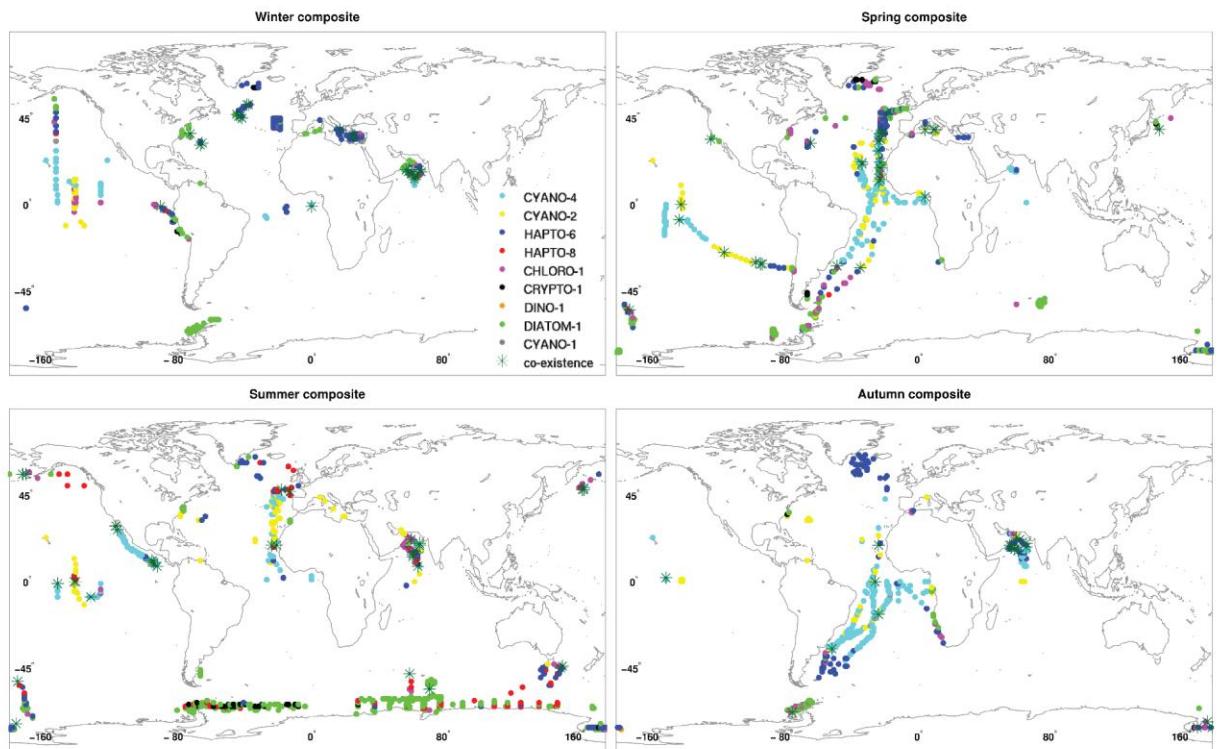


fig 7

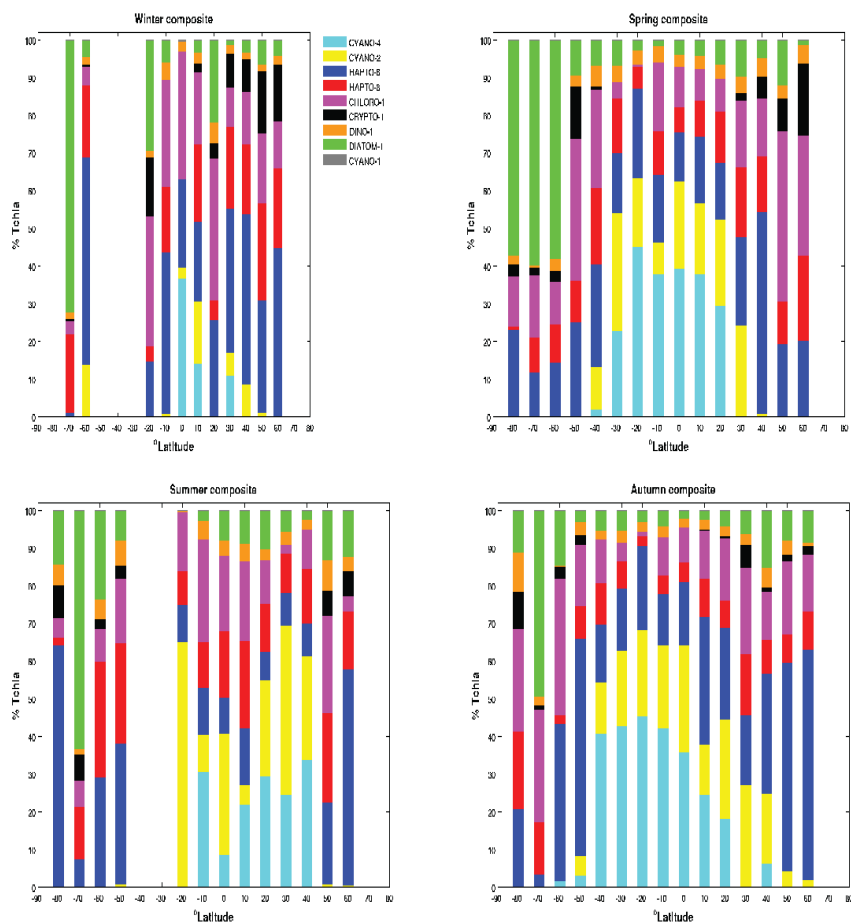


fig 8

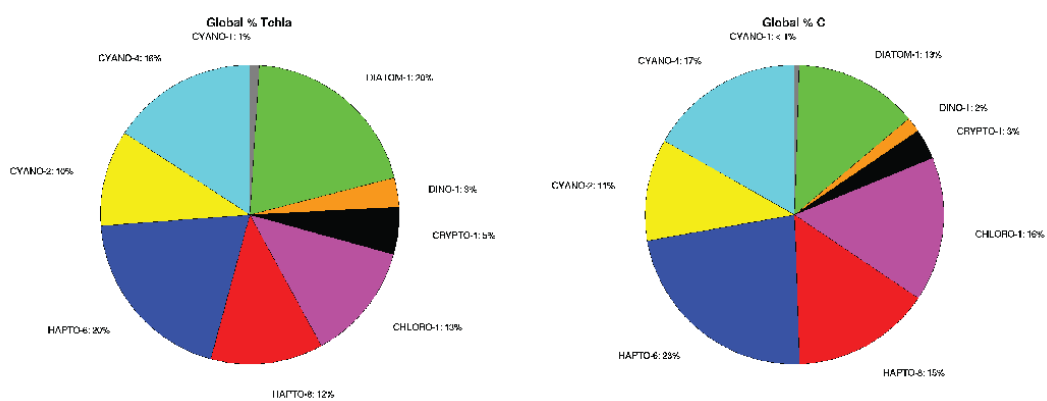


fig 9

Accepted manuscript