

The protist community mirrors seasonality and mesoscale hydrographic features in the oligotrophic Sargasso Sea

Leocadio Blanco-Bercial¹, Rachel Parsons¹, Luis Bolaños², Rod Johnson¹, Stephen Giovannoni³, and Ruth Curry¹

¹Bermuda Institute of Ocean Sciences

²University of Exeter

³Oregon State University

November 3, 2023

Abstract

Protists represent the majority of the eukaryotic diversity in the oceans. They have different functions in the marine food web, playing essential roles in the biogeochemical cycles. Meanwhile the available data is rich in horizontal and temporal coverage, little is known on their vertical structuring, particularly below the photic zone. The present study applies DNA metabarcoding to samples collected over three years in conjunction with the BATS time-series to assess marine protist communities in the epipelagic and mesopelagic zones. The protist community showed a dynamic seasonality in the epipelagic, responding to hydrographic yearly cycles. Mixotrophic lineages dominated throughout the year; however, autotrophs bloomed during the rapid transition between the winter mixing and the stratified summer, and heterotrophs had their peak at the end of summer, when the base of the thermocline reaches its deepest depth. Below the photic zone, the community, dominated by Rhizaria, is depth-stratified and relatively constant throughout the year, mirroring local hydrographic and biological features such as the oxygen minimum zone. The results suggest a dynamic partitioning of the water column, where the niche vertical position for each community changes throughout the year, likely depending on nutrient availability, the mixed layer depth, and other hydrographic features. Finally, the protist community closely followed mesoscale events (eddies), where the communities mirrored the hydrographic uplift, raising the deeper communities for hundreds of meters, and compressing the communities above.

Introduction

Marine protists encompass a large and heterogeneous community representing the majority of the eukaryotic diversity in the oceans (Worden *et al.* 2015). They include primary producers (autotrophs), heterotrophs (phagotrophs and parasitic), and a substantial collection of lineages exhibiting varying degrees of mixotrophic strategies (Mitra *et al.* 2016). These groups occupy distinct niches in the marine food web, playing essential roles in biogeochemical cycles (Mitra *et al.* 2014; Worden *et al.* 2015). They are responsible for ~50% of annual planktonic photosynthetic primary productivity (PP), of which they consume ~66%, plus an additional 10% of bacterial PP (Calbet & Landry 2004; Steinberg & Landry 2017). Recent applications of molecular tools such as metabarcoding (Choi *et al.* 2020; de Vargas *et al.* 2015), metagenomics and single cell genomics (Latorre *et al.* 2021), have significantly sharpened our understanding of protist diversity, distributions, and functionality, from basic trophic modes to complex metabolic pathways, and emphasized their importance in channeling marine productivity to upper trophic levels. The available observational data is rich in horizontal spatial and temporal coverage, yet lacks vertical resolution, particularly below the photic zone (Ollison *et al.* 2021). There, a large and diverse community of heterotrophic protists thrives on sinking particulate matter, preys upon the prokaryotic populations (Ollison *et al.* 2021; Rocke *et al.* 2015) and removes a similar

percentage of the prokaryotic standing stock compared to the epipelagic realm (Rocke *et al.* 2015). This trophic transfer may represent a critical mechanism sustaining the upper levels of mesopelagic trophic webs.

The warm oligotrophic subtropical gyres of the major ocean basins are the largest biome in the planet, and these nutrient-poor regions are expanding in size (Irwin & Oliver 2009). To predict future conditions, it is essential to characterize their present state (Agusti *et al.* 2019). In these regions, PP is dominated by the prokaryotic and eukaryotic picophytoplankton (Agusti *et al.* 2019; Cotti-Rauscher *et al.* 2020; Orsi *et al.* 2018; Riemann *et al.* 2011). Much of the PP of the larger size fractions depends on mixotrophic strategies, combining autotrophy with phagotrophy on the smaller primary producers (Mitra *et al.* 2016). The Bermuda Atlantic Time-series Study (BATS) is located in the western limits of the North Atlantic subtropical gyre (Lomas *et al.* 2013). The hydrography at BATS responds to a locally large seasonal cycle in atmospheric forcing, reflected in winter mixing that extends below the photic layer and contrasts sharply with a stratified summertime photic zone that is progressively mixed away during the fall. These characteristics drive nutrient availability, primary production and community composition (Church *et al.* 2013). Particle fluxes (Conte *et al.* 2001) and zooplankton community composition (Blanco-Bercial 2020) show a seasonal signal. Likewise, metazooplankton exhibit signals that reflect seasonal timescales, and it is expected that the protist community also exhibit seasonality, at least in the epipelagic layers.

A significant hurdle to studying the marine protist community is the complex and tedious procedure needed to characterize these organisms. Historically, this was achieved by microscopy, and required expert personnel for sample preparation, analyses and taxonomic identification. Microscopy or image-based methods have the advantage of being quantitative, providing an estimate of cell sizes, and enabling biomass proxies. These methods are limited, as any other morphology-based analyses, by the presence of cryptic species, and the natural complexity of the diversity of protists. In the oligotrophic ocean, a large proportion of the protist community falls within the 2-20 μm size class, many of them naked, which makes the identification very complicated when using automated optical instruments (e.g. FlowCam, IFCB), or even with classical methods after fixation. Molecular techniques as metabarcoding (Ollison *et al.* 2021) and metagenomics (Sunagawa *et al.* 2015) allow for direct characterization of planktonic communities (Not *et al.* 2007; Treusch *et al.* 2012; Treusch *et al.* 2009) and its potential functionality. These methods provide a better understanding of the diversity of the protist community in the oceans (Caron & Hu 2019) and the associated trophic relationships, including the large presence of symbionts (Ollison *et al.* 2021).

The present study applies DNA metabarcoding to samples collected over three years in conjunction with the BATS time-series to assess marine protist communities in the epipelagic and mesopelagic zones. Its purpose is to describe the vertical and temporal distributions of community composition, their evolution and transitions over the annual cycle, and relationships to hydrographic structure. We aimed to characterize the distinct communities with depth and detail how their boundaries evolve and transition, as a response to the hydrographic features throughout the year. In lieu of traditional approaches that bin data in depth layers, and seasons by calendar months, we employ a physical framework defined by four seasons and 11 vertical layers that delineate dynamical zones shaped by locally varying ocean mixing, stratification, and light penetration in the Sargasso Sea (Table 1; Curry *et al.*, in prep)

Methods

Oceanographic data

Monthly CTD profiles of temperature, salinity, density, oxygen, and chlorophyll fluorescence paired with discrete bottle sampling, provided hydrographic context. The details of sampling, quality control and post-processing are available on the BATS website (<http://bats.bios.edu/index.html>). Each profile was assigned to one of four seasons (mixed, spring, stratified, fall; Table 1) based on retrospective evaluation of the full time series for each year. Vertical layer boundaries (Table 1) were determined on a profile-by-profile basis from the fluorometer and density criteria.

Biological samples were acquired monthly from February 2016 to December 2018. March 2016 was sampled twice (March 8th and 25th); March 2017 sampling was cancelled due to rough weather. The protocol included

12 fixed depths between surface and 1000 m. At each depth, 4 L of water were filtered using 0.22 μ M Sterivex filters (MERCK, MA, USA). Filtering pressure was kept at 5 “ Hg. Both ends were sealed with parafilm or a sterile plug/cap. Filters were wrapped in foil and stored at -80°C.

Molecular methods

Sucrose lysis buffer (1mL) was added to the Sterivex filters upon thawing. DNA was extracted using the method by Giovannoni *et al.* (1990). Amplification and sequencing was carried out in the Center for Genome Research and Biocomputing at Oregon State University. The 18S rRNA V4 hypervariable region was amplified using the primers 5-CCAGCA[GC]C[CT]GCGGTAATTCC-3 and 5-ACTTTCGTTCTTGAT[CT][AG]A-3 (Stoeck *et al.* 2010), using the KAPA HiFi HotStart ReadyMix (Kapa Biosystems; Wilmington, MA, USA). The PCR thermal cycler program consisted of a 98°C denaturation step for 30 s, followed by 10 cycles of 10 s at 98°C, 30 s at 53°C and 30 s at 72°C, and 15 cycles of 10 s at 98°C, 30 s at 48°C and 30 s at 72°C, and a final elongation step at 72°C for 10 min. After PCR, dual indices and Illumina sequencing adapters were attached using the Illumina Nextera XT Index Kit. Amplicon sequencing was conducted on an Illumina MiSeq using 2x250 PE V2 chemistry.

Bioinformatics and analyses

Quality control and read merging were carried out with MeFiT (Parikh *et al.* 2016), using Jellyfish Ver. 2 (Marçais & Kingsford 2011), with strict parameters to compensate for the lack of complete overlap between forward and reverse: CASPER (Kwon *et al.* 2014) K -mer size of 5, threshold for difference of quality score 5, threshold for mismatching ratio 0.1, and minimum length of overlap of 15. The Merge and Quality Filter Parameters (using the meep score (Koparde *et al.* 2017)) were set at a filtering threshold of 0.1, discarding non-overlapping reads. Assembled reads passing the QC were fed into MOTHR ver. 1.43 (Schloss *et al.* 2009). In short, after trimming to the V4 region (by aligning to the SILVA 138 release trimmed to the V4 region) and removal of chimeras with VSEARCH (Rognes *et al.* 2016), single variants were obtained using UNOISE2 (Edgar 2016) as implemented in MOTHR (diffs=1). Taxonomic assignment was done using the PR² database (Guillou *et al.* 2013). After excluding metazoans and “Eukaryota_unclassified”, samples were rarefied to 20,000 reads and all ASVs with less than two reads (global singletons) were equaled to zero. Scripts are available at the GitHub repository https://github.com/blancoercial/Protist_Time_Series.

Counts per sample and environmental metadata were loaded into PRIMER Ver. 7 (Clarke & Gorley 2015). Diversity indices (S, H', J') were calculated in PRIMER; Faith's phylogenetic diversity PD (Faith 1992) was calculated in MOTHR using the same input table as for the other indices. Samples were then standardized by the total, and square-root transformed (Hellinger transformation) (Legendre & Gallagher 2001). A Bray-Curtis distance similarity matrix was built, and a Principal Coordinates Analysis (PCoA; Gower 1966) was carried out. The relationship between environmental variables and the similarity between samples was analyzed with BioEnv in PRIMER, using Spearman Rank and up to five variables for the model. The variables included day of year, depth, temperature, salinity, oxygen, fluorescence, turbidity and density. A second analysis included the factor “hydrographic layer” (Table 1), to assess how this partitioning compared to the protist community. All environmental variables were normalized (subtracting the mean and dividing by the standard deviation) before analyses to remove biases associated to different variable scales.

Non-hierarchical clustering based on group-average ranks was used to find the best grouping reflecting the self-organization of the community, from number of clusters $K = 2$ until two of the clusters were not statistically different ($K = 14$). At each K level, an analysis of similarities (ANOSIM) was run (with 999 permutations) to assess the R -value at each clustering level. The K levels showing the maximum R -values were further analyzed. Clustering was mapped to the hydrographic layers and evaluated within the PCoA ordination, and their taxonomic composition at PR² “Clade” level analyzed.

The molecular lineages (single variants) with the largest contribution to the similarity in each cluster, and to dissimilarity between adjacent clusters, were assessed using a Similarity Percentage (SIMPER) analyses. Fine taxonomic assignment was done using BLAST (Altschul *et al.* 1990), analyzing several of the top scores to understand the precision of assignments (species/genus/family/etc.). Interpretation the changes was

contextualize within the oceanographic landscape and transitions between clusters, and with the functional profile (based in the literature) of each lineage.

PCoA analyses were further performed at each discrete depth to study the interaction between seasonality and depth. Same procedure was repeated following those samples that overlapped with the Chl-a maximum peak for each profile (which did not happen every month because a set of fixed depths were sampled each time). SIMPER analyses were run to detect taxa related to seasonal succession at each depth.

For all analyses, a Holm-Bonferroni correction was applied to the significance threshold when multiple comparisons were done (Holm 1979).

Results

Hydrography

Over the 3-year period, hydrographic sampling revealed the expected seasonal cycles of cold season mixing, warm season stratification, the short spring transition, and the vertical distribution of chlorophyll fluorescence (Figure 1). The vertical layer boundaries fluctuated in depth over the observational period, highlighting the ability of this physical framework to transcend the temporal variations in vertical water mass structure associated with winter mixing, mesoscale eddies, location of the nutricline, and varying light penetration (Figure 1, Supp. Figure S1). The depth of the surface mixed layer (Layer 0) varied seasonally from 10 m in the stratified season to 170-212 meters in the mixed. The depth of the DCM (Deep Chlorophyll Maximum; Layer 2) ranged from 70 to 130 meters. The top of Layer 8, corresponding to the deep O₂ minimum, varied between 600 and 850 m. Other hydrographic features detected were the uplifts in the deep mesopelagic layers due to the passage large eddies (Figure 1). The most prominent uplifts (of about 200 m) occurred during 2018.

Molecular data

Of the 408 samples (34 sampling months, 12 depths), 369 produced a library. The remaining samples failed at some point of the procedure (sampling at sea or extraction), and did not produce a useable library. It usually affected several samples from the same cast, especially the deeper layers (where DNA concentration was always lower). About ~39 million reads were generated (average ~105,000 reads per sample; 23,000 to 364,000). Of those, the strict QC retained 42% as very high quality (16.5M reads; 44,000 per sample average). There was no link between reads passing QC and months or depths.

Vertical structure of communities

Non-hierarchical clustering based on Bray-Curtis distances showed two local R maxima, at $K = 2$ (ANOSIM $R = 0.903$; $p < 0.001$) and $K = 9$ ($R = 0.904$; $p < 0.001$). In both cases, all pairwise comparisons were significant. The $K = 2$ partitioning roughly divided the epipelagic (photic) and the mesopelagic zones, although when the MLD (Mixed Layer Depth) reached into in the mesopelagic, the upper community followed the MLD depth (Figure 2). The $K = 9$ clustering aligned closely with the hydrographic layering, with higher numbers of clusters in the epipelagic, and greater homogeneity in mesopelagic. In some cases, single communities spanned several mesopelagic layers. The clusters precisely traced hydrographic events such as the deepening of the MLD in winter, the abrupt succession to a different community following the shoaling of this layer at the spring transition, and the uplift of mesopelagic layers due to the passage of mesoscale eddies.

The epipelagic zone was characterized by a succession of different communities over the annual cycle. Cluster 1 dominated Layer 0 (surface MLD) in the spring and stratified seasons, but was displaced by Cluster 2 during the fall and mixed seasons. Cluster 1 was detected as soon as the MLD shoaled in each year, and remained detectable until the onset of Fall when the MLD (Layer 0) gradually deepened and Cluster 2 became dominant. Cluster 3, found below 1 and 2, corresponded to the lower portion of the epipelagic zone that included the DCM (Figure 2). Clusters 4, 5 and 6 occupied the underlying water mass (upper mesopelagic), corresponding to the local deep winter mixed layer (the Winter Mode Water, WMW; Layers 3 & 4), a very weakly stratified portion of the water column occupying the approximate depth range 150-400m, that did

not strictly align with density. Clusters 7, 8 and 9, in the lower mesopelagic (500 – 1000m), aligned with the density fields. Cluster 8 overlapped with Layer 8, the oxygen minimum zone (OMZ, with $O_2 < 160 \mu\text{mol kg}^{-1}$). The mesopelagic layers were uplifted during the passage of the fronts in 2018, detected by the raised density layers (Figures 2). Clusters were more tightly grouped with depth (PERMDISP test for homogeneity of multivariate dispersions; $p < 0.0001$), indicating a more variable community in the epipelagic compared to the mesopelagic, and in the upper mesopelagic compared to the deep mesopelagic.

The PCoA ordination arranged the samples into an arch, with the first axis roughly corresponding to depth (Figure 3). The BioEnv procedure determined that the best model combined O_2 , fluorescence, depth and density ($\rho=0.729$ $p < 0.001$), but since depth and density are highly correlated, the next best model included O_2 , fluorescence and density ($\rho=0.690$, $p < 0.001$). The proposed hydrographic layering was the best-correlated single environmental variable ($\rho=0.651$, $p < 0.001$). Superimposing the non-hierarchical clustering onto the PCoA, the $K=2$ showed a sharp divide between the clusters, despite the intrusions of the upper cluster into deeper layers. For $K=9$, separation of clusters 1 & 2 reflected the seasonal shifts in the upper epipelagic layers. The remaining clusters projected consecutively by depth/density in the PCoA, with very little mixing at the boundaries. The PCoA also indicated a larger separation between cluster 7 (occupying depths 500 – 600 m) and the other communities. In contrast, clusters 8 (OMZ) and 9 were very close to each other, with minimal mixing.

The taxonomic composition showed 40-45% of the reads corresponding to Syndiniales, a parasitic group (Figure 4); the proportion originated from free dispersal states versus the parasitic state is unknown. At $K=2$, the taxonomic composition showed a shift from a more diverse community in the epipelagic, with many different clades including autotrophs, heterotrophs and mixotrophs, to a heterotrophic, Rhizaria-dominated, community at depth (Figure 5). This result was more pronounced if the Syndiniales were not considered in the analyses. SIMPER analyses indicated that the epipelagic cluster was characterized by a large presence of autotrophs (mostly Pelagophyceae *Pelagomonas calceolata*) and mixotrophs (from Ochrophyta, Stramenopiles and Dinophyta) although some of the latter clades include all trophic possibilities. In contrast, heterotrophs overwhelmingly dominated the deep cluster, principally Radiolarians (Polycystinea and Acantharea), and representatives of Stramenopiles (Labyrinthulea). The dissimilarity between groups was then driven by the mixture of autotrophs (especially *P. calceolata* and the Chlorophyta *Ostreococcus* sp.) and heterotrophs/mixotrophs (Stramenopiles MAST-4A, and several Dinophyceae lineages) in the epipelagic, compared to their virtual absence in the mesopelagic, where depth-specific heterotrophic Radiolaria dominated.

The taxonomic composition at $K=9$ reflected a more detailed partitioning (Figure 4). Alveolata dominated the near-surface community during the spring and stratified seasons even if excluding Syndiniales (cluster 1), together with a mixture of Hacrobia and Stramenopiles. In contrast, the near-surface taxa during the mixed periods (cluster 2) exhibited greater abundances of Stramenopiles and Rhizaria. Cluster 3, below clusters 1 & 2, contained the highest concentrations of Archaeplastida, although Rhizaria was the dominant free-living group. A gradual decrease of all the non-Rhizaria groups was associated with increasing depth: Hacrobia and Stramenopiles disappeared almost completely by cluster 7, with only Alveolata maintaining a significant presence. Syndiniales abundances slightly decreased with depth, but had a greater presence in Cluster 7 (500-600 m). SIMPER analyses complimented these findings, indicating that taxa showed a clear affinity with cluster. There was no single ASV/OTU with a significant presence in all clusters, and most were significant only in one or two clusters. A few Rhizaria, however, showed high numbers in several mesopelagic layers (Figure 5; Table 1). Cluster 1 (surface stratified) was characterized by mixotrophs and heterotrophs of Alveolata Dinophyta (e.g., *Warnowia* sp., *Karlodinium* sp. or *Lepidodinium* spp.), Hacrobia Prymnesiophyceae (*Chrysochromulina* sp.) and several Stramenopiles. The only autotroph was a clade of *Phaeocystis* sp., although this group is known to appear as free living or as a symbiont autotroph in Rhizaria colonies. Cluster 2 (near-surface, mixed periods) shows large abundances of *P. calceolata* (autotroph), but most of the other clades belong to mixotrophs (e.g. several Dinophyceae, and MOCH-2), or heterotrophs (e.g., MAST lineages 25, 4A, 4C and 1D, Hacrobia Pterocystida). Cluster 3 (DCM) showed several mixotrophs among the dominant lineages, but the main groups were autotrophs (*P. calceolata*, *Ostreococcus* sp., *Bathycoccus*

prasinus, and the Haptophyta *Phaeocystis globosa*). Heterotrophs characteristic from this clade included the Hacrobia *Leucomyces marina* and several Rhizaria (although these might have autotrophic symbionts such as *Phaeocystis*). Clade 5, below the Chlorophyll maximum, showed significant abundances of the autotroph *P. calceolata*, however the shift towards a heterotrophic Rhizaria-dominated community was noticeable, together with heterotrophic (*Telonemia* sp., *Leucomyces marina*) and likely mixotrophic (Prymnesiophyceae Clade E) Hacrobia. Different Rhizaria dominated all remaining clusters, while some heterotrophic Hacrobia and Stramenopiles were still among the dominant clades in clusters 5 and 6.

The dissimilarity between consecutive clusters reflected transitions in community function. The mixotroph-dominated cluster one (surface stratified) was replaced by the autotroph-dominated community occupying the ML during the periods of deep mixing (cluster 2). In cluster 3 there was a larger increase of the autotrophs at the cost of mixotrophs from cluster 2. Deeper, differences were due to the progressive decrease in autotrophs, increase of heterotrophs, and layer to layer replacements between different heterotrophs (mostly Rhizaria).

Seasonality

When considering all samples together, there was no discernible seasonal signal. Seasonality, however, appeared when analyzing single depths/clusters. Seasonality was evident in the top layers (surface and 40 m especially; ANOSIM $R = 0.66$ and $R = 0.62$ respectively; $p < 0.001$), but the signal faded with depth, with ANOSIM R progressively decaying, and becoming non-significant below 300m. There was no significant seasonality in the mesopelagic. In the lower mesopelagic the passage of eddies was, however, detectable as a group of dissimilar samples (Figure 6).

Near the surface, the highly diverse community was composed of many taxa with low relative abundances (mostly $< 2\%$ average), dominated by mixotrophs and heterotroph clades. Only winter and spring had an autotroph (*P. calceolata*) among the most abundant clades (only representing $\sim 2\%$ of the reads), while a single heterotroph identified as *Warnowia* sp. represented 8% of reads in spring. By summer, mixotrophs, and a diverse community of heterotrophs, including *Telonemia*, *Warnowia*, several Hacrobia Centroheliozoa and the Rhizaria *Minorisa minutad* dominated the community. In the fall, mixotrophs gained more prominence compared to heterotrophs.

Cluster 3 was generally associated with the Layer 2 (broadly defined DCM) but, since the DCM varied substantially, sometimes this layer did not capture the feature. Cluster 3 community exhibited depth dependence, but no apparent seasonal cycle. When DCM samples were restricted to those acquired within 90% of the actual chlorophyll maximum, a strong and significant seasonal pattern was detected (ANOSIM $R = 0.0762$, $p < 0.001$; Figure 6). The stratified and fall periods, despite mixing in the 2PCos representation, were statistically different ($R = 0.5$, $p < 0.05$) and were separated in the 3-PCoA (Figure S2). SIMPER analyses on this narrow Chl-*a* max showed that the main difference is the spring season, during which a few autotrophic clades dominate the community (*Ostreococcus* sp., *P. calceolata* and *B. prasinus*). Three different clades of *Micromonas*, and *P. globosa*, were also among the most abundant clades. Together, autotrophs represented over 30% of the reads of the non-parasitic community. Other main groups included Dinophyceae (likely mixotrophs) and heterotrophs such as *Telonemia* sp. and Radiolaria (although these might have autotrophic endosymbionts). After transitioning to summer, only *P. calceolata*, *B. prasinus* and *P. globosa* showed relatively high abundances in the DCM; however, their prevalence was much lower (just above 5%). In contrast, mixotrophs and heterotrophs increased in relative abundance. In the fall transition, only *P. calceolata* and *P. globosa* were the autotrophs among the main taxa (with slightly higher relative abundances compared to summer), while the proportion of heterotrophs (Rhizaria and *Telonemia*) increased. During winter, there was again a strong shift to autotrophs (especially *P. calceolata*, *B. prasinus* and *Ostreococcus* sp.; about 10% of the total reads combined) and mixotrophs, with no free heterotrophs among the main lineages in the community.

Diversity

Diversity indices indicated a mismatch between species-based and phylogenetic-based diversity indices (Figure 7). The data showed a higher number of species (S), species diversity (H') and evenness (J') in the upper

layers of the water column, gradually decreasing with depth, reaching a minimum in the cluster corresponding to the oxygen minimum zone. These profiles also revealed an interesting feature: superimposed upon the general trend of decreasing diversity and evenness with increasing depth throughout the upper 1000m, the indices identified three subgroups (clusters 1-3, 4-6 and 7-9) each of which exhibited its own decreasing trend. The subgroups correspond respectively to the epipelagic, upper mesopelagic, and lower mesopelagic portions of the water column. The profile for Faith's phylogenetic diversity was distinctly different from these: it showed rising diversity from clusters 1 to 4, followed by a steady decrease with depth to a minimum in the OMZ.

Discussion

The data presented here supports the presence of a significant seasonality in the photic community in the Subtropical North Atlantic, whose signal and strength gradually decreases with depth, becoming undetectable in the mesopelagic. The structure of epipelagic and mesopelagic protist communities in the Sargasso Sea is strongly shaped by local ocean dynamics (e.g., mixing, stratification, nutrient distributions and light penetration) and expressed in the vertical and seasonal partitioning of community composition, diversity, and function. Mixotrophic groups dominate the sunlit epipelagic layers, with autotrophic taxa gaining relevance where the base of the photic layer meets the top of the nutricline (Arenovski *et al.* 1995). The mesopelagic layers below are inhabited by heterotrophic communities where Rhizaria becomes the most abundant group, outside of parasitic lineages (i.e. Syndiniales). The deep communities are likely fueled by complex food webs, feeding on prokaryotes, particles and each other (Byung Cheol *et al.* 2000; Rocke *et al.* 2015).

Nutrient limitation in the epipelagic zone favors mixo- and heterotrophic protists closer to the surface, with autotrophs balancing nutrients and light at its base. This is largely a consequence of ocean stratification, which enhances or inhibits vertical exchanges. The degree of stratification at the base of the photic zone affects the rate of nutrient supply fueling the autotrophic lineages at the local DCM. Over the annual cycle, hydrographic Layers 0, 1 and 2 exhibit dramatic changes in stratification driven by local air-sea fluxes, with Layers 1 and 2 being subsumed into Layer 0 beginning in the Fall and ending at the Spring transition. This process creates the vertical partitioning of clusters 1, 2 and 3 and their seasonality observed in K=9: i.e. clusters 1 and 2 dominate Layers 0 and 1 above the DCM from Spring through the Stratified season, with cluster 1 disappearing in the Fall and Mixed seasons, while cluster 3 in Layer 2 waxes in Spring then gradually wanes through the Stratified and Fall seasons. Such seasonality is absent at ALOHA in the North Pacific where winter mixing is shallower, the base of the photic layer remains stratified throughout the year, and protist communities align with depth (Ollison *et al.* 2021). This difference is not unexpected, due to the more constant hydrographic conditions, primary production and particle flux measured throughout the year by the Hawaii Ocean Time-series compared to those at BATS (Church *et al.* 2013). The more homogeneous conditions would then favor the presence of a dominant community all year long at ALOHA, showing only depth-controlled layering.

The pronounced seasonality observed in the Sargasso Sea implies shifts in trophic ecology and energy fluxes which likely influence biogeochemical cycles over the annual cycle. Elevated abundances of mixotrophs and heterotrophs relative to autotrophs during the stratified and fall seasons indicate a complex recycling food web where small protists such as *Warnowia*, *Telonemia*, *Karlodinium* or *Minorisa minuta* and MAST and MOCH lineages prey on picophytoplankton, responsible of most of the primary production in the Sargasso Sea (Caron *et al.* 1999; Cotti-Rausch *et al.* 2020; del Campo *et al.* 2013; Klaveness *et al.* 2005; Massana *et al.* 2014; Orsi *et al.* 2018; Place *et al.* 2012; Riemann *et al.* 2011; Sanders *et al.* 2000). Both autotrophs and small mixo- and heterotrophs would also be preyed upon by larger mixotrophs or heterotrophs protists (Andersen *et al.* 2011; Evelyn & Michael 1998; Quevedo & Anadón 2001). Autotrophic protists (picophytoplankton and small nanophytoplankton, such as *Ostreococcus*, *Bathycoccus* or *Pelagomonas calceolata*) had a numerical relevance only in the mixed and spring seasons, but mixotrophs still represent a significant portion of the community. Those mixotrophs would benefit from their ability to carry out photosynthesis, while still preying on the small autotrophs (Sanders *et al.* 2000), likely due to the concurring C fixation by photosynthesis with the

heterotrophic acquisition of N and P (Edwards 2019). Recently, Choi *et al.* (2020) reported elevated summer abundances of uncultured dictyophytes at BATS, and evidence that these taxa are more prevalent in the upper water common when nutrients are most depleted. These and other observations, and the findings we report, suggest that mixotrophic strategies become relatively more advantageous when production is limited by macronutrients, likely due to the repartitioning of N and P rather than the energetic benefits of mixotrophy. As such, in the oligotrophic Sargasso Sea they would be at a significant advantage compared to non-mixotrophs (Choi *et al.* 2020; Duhamel *et al.* 2019; Edwards 2019), only attenuated during the strong mixing period.

The observed month-to-month depth expansion and contraction of the different protist communities influences community functionality. Cluster 1, which occupies the near surface layers throughout the spring and stratified season, disappears in fall when the surface MLD begins to erode into the DCM, and only emerges again when the ML abruptly shoals the following spring. The temporary nature of this group contrasts with the other communities, which are likely present throughout the year, albeit with periods of expansion and contraction. As such, heterotrophy, and its trophic and biogeochemical consequences, would increase in summer and fall, with the expansion of the stratified surface community. The fall is also when the Chl-*a* -max corresponds to a community rich in heterotrophs, although many are potential mixotrophs via endosymbionts (Bjorbækmo *et al.* 2020). The effect of these non-constitutive mixotrophs on the biogeochemical cycles is one major question to be addressed in the future. Although in some cases a few clusters are missing in some months, it might be an artifact of the punctuated depth sampling, which would miss their depths at that particular month. As a comparison, the bottle sampling also missed some of the density-derived layers, but they were all present in the continuous profiles. In those months, some of the deeper clusters were then missed.

It should be noted that in some cases the functionality of taxa is relatively easy to assign (e.g. *Telonemia* or *Ostreococcus*), while in many other cases it is difficult to assess with certainty. Close relatives, with different functional profiles (i.e. mixotrophy vs. autotrophy) often share a single V4 sequence. Other sources of uncertainty come from the non-constitutive mixotrophs. In our case, the main group is Rhizaria. These are known to potentially carry photosynthetic endosymbionts (e.g., *Phaeocystis* sp.) (Bjorbækmo *et al.* 2020). Addressing these uncertainties is required to understand the full role of mixotrophy in the oceans, and its influence in the biogeochemical cycles (Edwards 2019; Flynn *et al.* 2019; Gonçalves Leles *et al.* 2021; Gonçalves Leles *et al.* 2018; Mitra *et al.* 2014).

In the mesopelagic zone, protist are vertically partitioned into distinct communities that occupy specific density strata. This may reflect a response to resource availability and/or particular environmental conditions (Biard & Ohman 2020), such as the OMZ, where there is an increase in particles. Within the main clades, specific lineages also occupied specific vertical layers, suggesting a preferred niche for each. An alternate explanation is that vertical biological gradients may reflect “where” and “when” the density layers were at the sea surface – i.e. the ventilated thermocline model of gyre circulation (Luyten *et al.* 1983). The “age” of a water parcel at BATS typically increases with depth/density reflecting longer distances traveled since ventilation. Geographic gradients in surface community composition could thus translate to vertical gradients at mesopelagic depths, and/or, if the community composition evolves as a function of time, the vertical gradients may reflect “age”.

Diversity patterns

Our findings, showing higher diversity in the epipelagic, peaking below the Chl-*a* max., and then decreasing with depth, are opposite in pattern to the study by Ollison *et al.* (2021). These differences can be attributed to many causes. There is great disparity in the environmental conditions, with strong seasonality and fully oxic OMZ at BATS compared to the quasi-permanent conditions and the thick and suboxic OMZ at ALOHA (Church *et al.* 2013) . Additionally there are pure methodological differences: the present manuscript is based in DNA metabarcoding, while Ollison *et al.* (Ollison *et al.* 2021) used an RNA-based approach. Our data, however, coincides with the findings of Canals *et al.* (2020), where a similar pattern of decreasing diversity with depth and a peak around the Chl-*a* maximum, was found for ciliates in the Atlantic Ocean, suggesting

that the environmental landscape (the hydrography) is driving the differences between ocean basins.

The clear signal of increasing phylogenetic diversity towards the Chla-max, with the layers below the 1% (but above the mesopelagic) showing the highest diversity, is notable. This maxima might be related to the presence of more, and more diverse, ecological niches; here light and nutrients are able to sustain a rich autotroph community (below it is too dark; nutrients are scarce above). This pattern contrasts with the taxonomic diversity, either H' or SVs richness, which did not show a noticeable change in the epipelagic. This mismatch is significant since PD is often directly correlated to species richness (Barker 2008; Voskamp *et al.* 2020). The pattern here likely reflects the abrupt transition from an Alveolata-dominated surface community towards one more phylogenetically diverse, despite a relatively constant species richness, owing to the availability of nutrients. The overlapping presence of light and nutrient gradients would favor the presence of more, and more diverse, ecological niches, facilitating the coexistence of more phylogenetically apart lineages.

The mesopelagic communities exhibited lower diversity and tighter grouping in the PCoA, indicative of less month-to-month variability compared to epipelagic communities. These patterns may be related to seasonality in the epipelagic layers, where a more dynamic environment would favor community succession processes month to month (represented as a higher dispersal between samples) and inhibits efficient competitive exclusion processes (causing higher diversity within sample). The deeper layers, however, characterized by less environmental variability, would favor the same lineages with competitive advantages month after month, limiting the number of taxa, and resulting in more stable communities.

Methodological caveats

Metabarcoding approaches, as any other method, are subject to many methodological limitations and bias (Bucklin *et al.* 2016; Santoferrara 2019). The initial selection of which region and primers to use, the marker copy number variability between taxa, etc., are a source of biases that would affect the description of the community. Other source of uncertainty is the taxonomic assignment of the different lineages (SV or OTUs), due to both incomplete databases (where there is a lot of work to do by the community), and to the lack of resolution at the species level by most ribosomal markers. To mitigate these biases, the taxonomy for the discussed lineages was analyzed against GenBank, to understand the ambiguity of the taxonomic assignment, and corrected if needed (compared to the mothur PR² based assignment). Regarding the copy number bias, we are confident that the core of this manuscript, the seasonal and depth community transitions, are not affected by this fact, and would very likely hold true independently of the marker used. Finally, diversity indices (the raw value itself) should be considered only in the context of this study, since methodological questions as the indicated or the rarefaction (or extrapolation) level chosen can drastically affect diversity estimates (Blanco-Bercial 2020; Santoferrara 2019). Again, despite any possible bias, it is likely that the observed patterns of variability (where diversity is higher/lower, etc.) are robust against those bias. Any comparison with other existing data would require a reanalysis with shared bioinformatics approaches.

Conclusion

The protist community in the Sargasso Sea shows a dynamic seasonality in the epipelagic, responding to hydrographic yearly cycles. Mixotrophic lineages, able to take advantage of the smaller picophytoplankton and heterotroph bacterioplankton, dominate throughout the year; however, autotrophs bloom during the rapid transition between the winter mixing and the stratified summer. Pure heterotrophs have their peak moment at the end of summer, when the base of the thermocline reaches its deepest depth, and likely mixotrophs lose part of their photosynthetic advantage. Below the photic zone, the community, dominated by Rhizaria, is depth-stratified, relatively constant throughout the year. These populations, relying on the vertical flux of organic matter, respond to local hydrographic and biological features such as the oxygen minimum zone. Together, this suggests a dynamic partitioning of the water column, where the niche vertical position for each community changes throughout the year, likely depending on nutrient availability, the mixed layer depth, and other hydrographic features. Future research should address several uncertainties needed to implement these data into biogeochemical models. The most prominent problem to solve is the lack of lineage-level quantitative data (i.e., counts), which would allow a better understanding of the energy

flows. Most of the lineages we report on here are 2-20 μm in cell diameter, falling in the range where taxon-specific cell counting is most difficult, between flow cytometry and automated image analyses. A second uncertainty is clarifying the functionality (trophic modes) in those lineages (e.g. MOCH groups with autotrophs, mixotrophs, and heterotrophs), to more accurately associate these strategies with the temporal and spatial patterns in the system. Together with the present data, these would allow better characterization of the carbon flux throughout the water column, and the role of mixotrophs and heterotrophs in controlling the biogeochemical cycles in the oligotrophic oceans.

Acknowledgements

This research was funded by the Simons Foundation International as part of the BIOSCOPE (<http://scope.bios.edu/>) and MAGIC (SFI award #545298) projects, the NSF funded OCE-1851224, and the Bermuda Atlantic Time-series Study program (BATS; NSF OCE-1756105). Special thanks to the BATS personnel and to the Captains and crew of the RV Atlantic Explorer for their logistical support at sea. This research made use of the FCMMF laboratory equipment at BIOS (NSF DBI-1522206).

Data Accessibility and Benefit-Sharing Statement

Raw sequence reads and metadata are deposited in the Bioproject PRJNA769790. Scripts are available at the GitHub repository https://github.com/blancoercial/Protist_Time_Series.

References Cited

- Agusti S, Lubián LM, Moreno-Ostos E, Estrada M, Duarte CM (2019) Projected Changes in Photosynthetic Picoplankton in a Warmer Subtropical Ocean. *Frontiers in Marine Science* **5** .
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215** , 403-410.
- Andersen NG, Nielsen TG, Jakobsen HH, Munk P, Riemann L (2011) Distribution and production of plankton communities in the subtropical convergence zone of the Sargasso Sea. II. Protozooplankton and copepods. *Marine Ecology Progress Series* **426** , 71-86.
- Arenovski AL, Lim EL, Caron DA (1995) Mixotrophic nanoplankton in oligotrophic surface waters of the Sargasso Sea may employ phagotrophy to obtain major nutrients. *Journal of Plankton Research* **17** , 801-820.
- Barker GM (2008) Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation. *Biological Journal of the Linnean Society* **76** , 165-194.
- Biard T, Ohman MD (2020) Vertical niche definition of test-bearing protists (Rhizaria) into the twilight zone revealed by in situ imaging. *Limnology and Oceanography* **65** , 2583-2602.
- Bjorbækmo MFM, Evenstad A, Røsæg LL, Krabberød AK, Logares R (2020) The planktonic protist interactome: where do we stand after a century of research? *The ISME Journal* **14** , 544-559.
- Blanco-Bercial L (2020) Metabarcoding Analyses and Seasonality of the Zooplankton Community at BATS. *Frontiers in Marine Science* **7** .
- Bucklin A, Lindeque PK, Rodriguez-Ezpeleta N, Albaina A, Lehtiniemi M (2016) Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *Journal of Plankton Research* **38** , 393-400.
- Byung Cheol C, Sang Cheol N, Dong Han C (2000) Active ingestion of fluorescently labeled bacteria by mesopelagic heterotrophic nanoflagellates in the East Sea, Korea. *Marine Ecology Progress Series* **206** , 23-32.
- Calbet A, Landry MR (2004) Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnology and Oceanography* **49** , 51-57.
- Canals O, Obiol A, Muhovic I, Vaqué D, Massana R (2020) Ciliate diversity and distribution across horizontal and vertical scales in the open ocean. *Molecular Ecology* **29** , 2824-2839.

- Caron DA, Hu SK (2019) Are We Overestimating Protistan Diversity in Nature? *Trends in Microbiology* **27** , 197-205.
- Caron DA, Peele ER, Lim EL, Dennett MR (1999) Picoplankton and nanoplankton and their trophic coupling in surface waters of the Sargasso Sea south of Bermuda. *Limnology and Oceanography* **44** , 259-272.
- Choi CJ, Jimenez V, Needham DM, *et al.* (2020) Seasonal and Geographical Transitions in Eukaryotic Phytoplankton Community Structure in the Atlantic and Pacific Oceans. *Frontiers in Microbiology* **11** .
- Church MJ, Lomas MW, Muller-Karger F (2013) Sea change: Charting the course for biogeochemical ocean time-series research in a new millennium. *Deep Sea Research Part II: Topical Studies in Oceanography* **93** , 2-15.
- Clarke KR, Gorley RN (2015) PRIMER v7: User Manual/Tutorial. *PRIMER-E, Plymouth* , 296.
- Conte MH, Ralph N, Ross EH (2001) Seasonal and interannual variability in deep ocean particle fluxes at the Oceanic Flux Program (OFP)/Bermuda Atlantic Time Series (BATS) site in the western Sargasso Sea near Bermuda. *Deep Sea Research Part II: Topical Studies in Oceanography* **48** , 1471-1505.
- Cotti-Rausch BE, Lomas MW, Lachenmyer EM, Baumann EG, Richardson TL (2020) Size-fractionated biomass and primary productivity of Sargasso Sea phytoplankton. *Deep Sea Research Part I: Oceanographic Research Papers* **156** , 103141.
- de Vargas C, Audic S, Henry N, *et al.* (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* **348** .
- del Campo J, Not F, Forn I, Sieracki ME, Massana R (2013) Taming the smallest predators of the oceans. *The ISME Journal* **7** , 351-358.
- Duhamel S, Kim E, Sprung B, Anderson OR (2019) Small pigmented eukaryotes play a major role in carbon cycling in the P-depleted western subtropical North Atlantic, which may be supported by mixotrophy. *Limnology and Oceanography* **64** , 2424-2440.
- Edgar RC (2016) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv* , 081257.
- Edwards KF (2019) Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proceedings of the National Academy of Sciences* **116** , 6211-6220.
- Evelyn JL, Michael CM (1998) Microzooplankton herbivory and phytoplankton growth in the northwestern Sargasso Sea. *Aquatic Microbial Ecology* **16** , 173-188.
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61** , 1-10.
- Flynn KJ, Mitra A, Anestis K, *et al.* (2019) Mixotrophic protists and a new paradigm for marine ecology: where does plankton research go now? *Journal of Plankton Research* **41** , 375-391.
- Giovannoni SJ, DeLong DF, Schmidt TM, Pace NR (1990) Tangentail flow filtration and preliminary phylogenetic analysis of marine picoplankton. *Applied and Environmental Microbiology* **56** , 2572-2575.
- Gonçalves Leles S, Bruggeman J, Polimene L, *et al.* (2021) Differences in physiology explain succession of mixoplankton functional types and affect carbon fluxes in temperate seas. *Progress in Oceanography* **190** , 102481.
- Gonçalves Leles S, Polimene L, Bruggeman J, *et al.* (2018) Modelling mixotrophic functional diversity and implications for ecosystem function. *Journal of Plankton Research* **40** , 627-642.
- Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53** , 325-338.

- Guillou L, Bachar D, Audic S, *et al.* (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research* **41** , D597-D604.
- Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6** , 65-70.
- Irwin AJ, Oliver MJ (2009) Are ocean deserts getting larger? *Geophysical Research Letters* **36** .
- Klaveness D, Shalchian-Tabrizi K, Thomsen HA, Eikrem W, Jakobsen KS (2005) *Telonema antarcticum* sp. nov., a common marine phagotrophic flagellate. *International Journal of Systematic and Evolutionary Microbiology* **55** , 2595-2604.
- Koparde VN, Parikh HI, Bradley SP, Sheth NU (2017) MEEPTOOLS: a maximum expected error based FASTQ read filtering and trimming toolkit. *International Journal of Computational Biology and Drug Design* **10** , 237-247.
- Kwon S, Lee B, Yoon S (2014) CASPER: context-aware scheme for paired-end reads from high-throughput amplicon sequencing. *BMC Bioinformatics* **15** , S10.
- Latorre F, Deutschmann IM, Labarre A, *et al.* (2021) Niche adaptation promoted the evolutionary diversification of tiny ocean predators. *Proceedings of the National Academy of Sciences* **118** , e2020955118.
- Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* **129** , 271-280.
- Lomas MW, Bates NR, Johnson RJ, *et al.* (2013) Two decades and counting: 24-years of sustained open ocean biogeochemical measurements in the Sargasso Sea. *Deep Sea Research Part II: Topical Studies in Oceanography* **93** , 16-32.
- Luyten J, Pedlosky J, Stommel H (1983) The ventilated thermocline. *Journal of Physical Oceanography* **13** , 292-309.
- Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27** , 764-770.
- Massana R, del Campo J, Sieracki ME, Audic S, Logares R (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *The ISME Journal* **8** , 854-866.
- Mitra A, Flynn KJ, Burkholder JM, *et al.* (2014) The role of mixotrophic protists in the biological carbon pump. *Biogeosciences* **11** , 995-1005.
- Mitra A, Flynn KJ, Tillmann U, *et al.* (2016) Defining Planktonic Protist Functional Groups on Mechanisms for Energy and Nutrient Acquisition: Incorporation of Diverse Mixotrophic Strategies. *Protist* **167** , 106-120.
- Not F, Gausling R, Azam F, Heidelberg JF, Worden AZ (2007) Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environmental Microbiology* **9** , 1233-1252.
- Ollison GA, Hu SK, Mesrop LY, DeLong EF, Caron DA (2021) Come rain or shine: Depth not season shapes the active protistan community at station ALOHA in the North Pacific Subtropical Gyre. *Deep Sea Research Part I: Oceanographic Research Papers* **170** , 103494.
- Orsi WD, Wilken S, del Campo J, *et al.* (2018) Identifying protist consumers of photosynthetic picoeukaryotes in the surface ocean using stable isotope probing. *Environmental Microbiology* **20** , 815-827.
- Parikh HI, Koparde VN, Bradley SP, Buck GA, Sheth NU (2016) MeFiT: merging and filtering tool for illumina paired-end reads for 16S rRNA amplicon sequencing. *BMC Bioinformatics* **17** , 491.
- Place AR, Bowers HA, Bachvaroff TR, *et al.* (2012) *Karlodinium veneficum*—The little dinoflagellate with a big bite. *Harmful Algae* **14** , 179-195.

Quevedo M, Anadón R (2001) Protist control of phytoplankton growth in the subtropical north-east Atlantic. *Marine Ecology Progress Series* **221** , 29-38.

Riemann L, Nielsen TG, Kragh T, *et al.* (2011) Distribution and production of plankton communities in the subtropical convergence zone of the Sargasso Sea. I. Phytoplankton and bacterioplankton. *Marine Ecology Progress Series* **426** , 57-70.

Rocke E, Pachiadaki MG, Cobban A, Kujawinski EB, Edgcomb VP (2015) Protist Community Grazing on Prokaryotic Prey in Deep Ocean Water Masses. *PLoS ONE* **10** , e0124505.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4** , e2584.

Sanders RW, Berninger UG, Lim EL, Kemp PF, Caron DA (2000) Heterotrophic and mixotrophic nano-plankton predation on picoplankton in the Sargasso Sea and on Georges Bank. *Marine Ecology Progress Series* **192** , 103-118.

Santoferrara LF (2019) Current practice in plankton metabarcoding: optimization and error management. *Journal of Plankton Research* **41** , 571-582.

Schloss PD, Westcott SL, Ryabin T, *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75** , 7537-7541.

Steinberg DK, Landry MR (2017) Zooplankton and the Ocean Carbon Cycle. *Annual Review of Marine Science* **9** , 413-444.

Stoeck T, Bass D, Nebel M, *et al.* (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology* **19** , 21-31.

Sunagawa S, Coelho LP, Chaffron S, *et al.* (2015) Structure and function of the global ocean microbiome. *Science* **348** .

Treusch AH, Demir-Hilton E, Vergin KL, *et al.* (2012) Phytoplankton distribution patterns in the northwestern Sargasso Sea revealed by small subunit rRNA genes from plastids. *The ISME Journal* **6** , 481-492.

Treusch AH, Vergin KL, Finlay LA, *et al.* (2009) Seasonality and vertical structure of an ocean gyre microbial community. *ISME Journal* **3** , 1148-1163.

Voskamp A, Hof C, Biber MF, *et al.* (2020) Climate change impacts on the phylogenetic diversity of the world's terrestrial birds: more than species numbers. *bioRxiv* , 2020.2012.2002.378216.

Worden AZ, Follows MJ, Giovannoni SJ, *et al.* (2015) Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347** , 1257594.

Table 1. Hydrographic envelope for the vertical layers and season definitions. Vertical layers are defined based in density and the CTD-measured fluorescence. Seasons are defined based on the depth of the mixed layer (MLD), and the relative position of the deep Chl *a* -max (DCM) with it.

Vertical layer	Name	Definition	Definition	Definition
0	Surface Mixed Layer (ML)	0 m to base of surface mixed layer (MLD, defined as density $\geq \sigma_{\theta} \text{ surp} \alpha_{se} + 0.125 \text{ kg m}^{-3}$	0 m to base of surface mixed layer (MLD, defined as density $\geq \sigma_{\theta} \text{ surp} \alpha_{se} + 0.125 \text{ kg m}^{-3}$	0 m to base of surface mixed layer (MLD, defined as density $\geq \sigma_{\theta} \text{ surp} \alpha_{se} + 0.125 \text{ kg m}^{-3}$
1	Upper Stratified Euphotic Zone	MLD to top of DCM layer	MLD to top of DCM layer	MLD to top of DCM layer

Vertical layer	Name	Definition	Definition	Definition
2	Deep Chlorophyll Maximum (DCM)	Portion of Chl-a profile ≥ 0.35 of Chl-a max	Portion of Chl-a profile ≥ 0.35 of Chl-a max	Portion of Chl-a profile ≥ 0.35 of Chl-a max
3	Winter Mode Water layer	Base of DCM to density of local winter MLD ($\sigma_\theta = 26.32$)	Base of DCM to density of local winter MLD ($\sigma_\theta = 26.32$)	Base of DCM to density of local winter MLD ($\sigma_\theta = 26.32$)
4	Ventilated Thermocline Layer 4	Base of WMW to $\sigma_\theta 26.50$	Base of WMW to $\sigma_\theta 26.50$	Base of WMW to $\sigma_\theta 26.50$
5	5	$\sigma_\theta 26.50$ to 26.70	$\sigma_\theta 26.50$ to 26.70	$\sigma_\theta 26.50$ to 26.70
6	6	$\sigma_\theta 26.70$ to 26.90	$\sigma_\theta 26.70$ to 26.90	$\sigma_\theta 26.70$ to 26.90
7	7	$\sigma_\theta 26.90$ to 27.10	$\sigma_\theta 26.90$ to 27.10	$\sigma_\theta 26.90$ to 27.10
8	8 (Deep O ₂ minimum)	$\sigma_\theta 27.10$ to 27.30	$\sigma_\theta 27.10$ to 27.30	$\sigma_\theta 27.10$ to 27.30
9	9	$\sigma_\theta 27.30$ to 27.50	$\sigma_\theta 27.30$ to 27.50	$\sigma_\theta 27.30$ to 27.50
10	10	$\sigma_\theta 27.50$ to 27.90	$\sigma_\theta 27.50$ to 27.90	$\sigma_\theta 27.50$ to 27.90
Season	Definition	Year 1 (ends)	Year 2 (ends)	Year 3 (ends)
Mixed	MLD > DCM layer	01-Apr-2016	10-Apr-2017	05-Apr-2018
Spring	MLD abruptly shoals to ~100m	31-May-2016	26-Apr-2017	26-Apr-2018
Stratified	MLD shallower than top of DCM	01-Oct-2016	16-Oct-2017	01-Oct-2018
Fall	First entrainment of DCM into ML	01-Dec-2016	01-Jan-2018	01-Jan-2019

Figure 1. Density (σ_θ , kg m⁻³), oxygen ($\mu\text{mol kg}^{-1}$) and fluorescence (RFU) profiles at BATS station from March 2016 to December 2018. Thin, vertical lines indicate sampling events. Density and oxygen profiles are 0-1000 m, meanwhile fluorescence is shown only in the epipelagic (0-200 m). Mixed Layer Depth (MLD) is indicated by a white line in the density profile. The oxygen minimum zone is defined as the region with less than 160 $\mu\text{mol kg}^{-1}$, delineated with a black line in the oxygen profile.

Figure 2. Assignment for each sampled depth, at each month, to the community-composition based on non-hierarchical clusters for $K = 2$ (upper panel) and $K = 9$ (middle panel). $K = 2$ separates the epipelagic and the mesopelagic, meanwhile $K = 9$ reflects the epipelagic seasonality and depth-stratification. The lower panel shows the assignment of each sample to the 11 hydrographic layers defined by Curry et al. (see text). Blue bars correspond to deep mixing events, black bars to uplift of the deeper layers due to the passage of eddies.

Figure 3. PCoA ordination of all samples for the first two axis, based in their community composition. Environmental factors over imposed as vectors, with length corresponding to ρ . The circle represents $\rho = 1$. Colors indicate the non-hierarchical cluster for each sample, for $K = 2$ (upper panel) and $K = 9$ (lower panel).

Figure 4. Taxonomic composition (at clade level) for each non-hierarchical cluster at $K = 2$ (top) and $K = 9$ (bottom). The parasitic order Syndiniales is highlighted within the Alveolata, showing its dominance within this phylum.

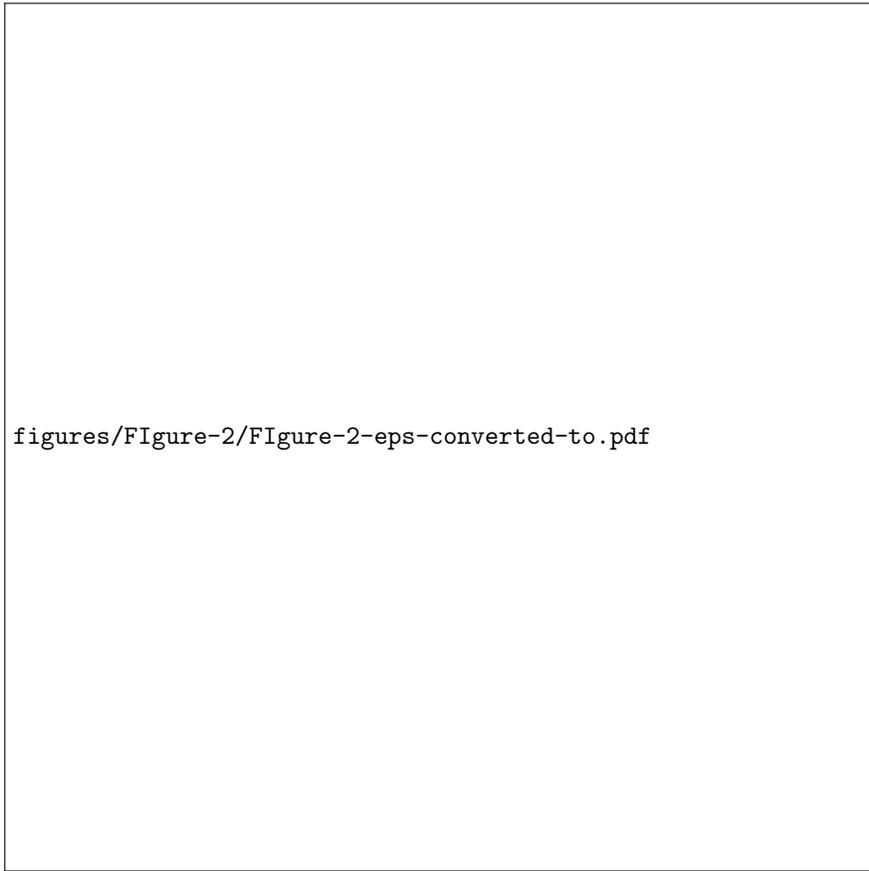
Figure 5. Functional profile (red: heterotroph, orange: mixotroph, green: autotroph) for each of the main components of the nine clusters, as indicated by the SIMPER analyses (after removal of parasitic lineages). When the V4 region was shared among lineages with different profiles, the corresponding colors are indicated in the bar.

Figure 6. PCoA ordinations for samples from surface (upper left), Chl-*a* maximum (upper right), 600 m depth (lower left) and 800 m depth (lower right), with ANOSIM R values for factor “seasons” indicated for each graph, and colors indicating seasons. In the epipelagic, a seasonal signal was detected, especially in the upper layers. In the mesopelagic (lower panels) no seasonality was observed, and samples diverging from the center core at these depths corresponded however to the uplifting events due to the passage of hydrographic features (within the squares).

Figure 7. Box and whiskers graphs for the distribution of the number of taxa (S), Shannon’s H, Faith PD and Pielou’s J for all samples from each cluster at $K = 9$. Boxes indicate median, first and third quartiles.

Hosted file

Figure 1.eps available at <https://authorea.com/users/695568/articles/551657-the-protist-community-mirrors-seasonality-and-mesoscale-hydrographic-features-in-the-oligotrophic-sargasso-sea>



figures/FIGure-2/FIGure-2-eps-converted-to.pdf

figures/Figure-3/Figure-3-eps-converted-to.pdf

figures/Figure-4/Figure-4-eps-converted-to.pdf

figures/Figure-5/Figure-5-eps-converted-to.pdf

figures/Figure-6/Figure-6-eps-converted-to.pdf

figures/Figure-7/Figure-7-eps-converted-to.pdf